



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

OpenStreetMap: Uma Análise Sobre a Evolução de Dados Geográficos Colaborativos no Brasil

Gabriel Franklin Braz de Medeiros

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Orientadora
Prof.^a Dr.^a Maristela Terto de Holanda

Brasília
2017



OpenStreetMap: Uma Análise Sobre a Evolução de Dados Geográficos Colaborativos no Brasil

Monografia apresentada como requisito parcial
para conclusão do Curso de Engenharia da Computação

Prof.^a Dr.^a Aletéia Patrícia Favacho de Araújo Prof. Dr. Márcio de Carvalho Victorino
CIC/UnB CIC/UnB

Prof. Dr. Ricardo Pezzuol Jacobi
Coordenador do Curso de Engenharia da Computação

Brasília, 21 de julho de 2017

Dedicatória

Dedico este trabalho aos meus familiares e amigos, os quais sempre estiveram presentes ao longo de minha vida e, em especial, àqueles que estiveram ao meu lado durante os momentos mais difíceis de minha caminhada acadêmica.

Agradecimentos

Agradeço primeiramente a Deus, por ter conseguido chegar até aqui. Agradeço aos meus pais, por todo o apoio fornecido durante a minha vida e por terem constituído a base da minha educação. Agradeço aos meus colegas e professores da Universidade de Brasília (UnB), fundamentais para o meu amadurecimento acadêmico. Agradeço ao pessoal do apoio tecnológico, do Departamento de Ciência da Computação (CIC) da UnB, por fornecer o *hardware* necessário para que os testes fossem realizados. Agradeço também à professora Maristela, orientadora e incentivadora deste trabalho.

Resumo

O OpenStreetMap (OSM) é uma ferramenta de mapeamento colaborativo desenvolvida no ano de 2004 pelo estudante de computação Steve Coast, da *University College London (UCL)*. Inicialmente, a ideia do projeto era realizar o mapeamento apenas da região do Reino Unido, mas logo despertou o interesse de pesquisadores de outros países. Uma vez que os usuários participam ativamente dos processos de inclusão, alteração e exclusão dos dados disponibilizados na ferramenta, é necessário que exista uma constante preocupação em relação à qualidade e consistência dos mesmos. Dessa forma, o objetivo principal deste trabalho é realizar uma análise acerca da evolução dos dados geográficos do Brasil inseridos dentro dessa plataforma, tomando como base o período compreendido entre os anos de 2007 e 2016.

Palavras-chave: OpenStreetMap, Dados Geográficos, Sistemas de Informações Geográficas, Crowdsourcing, Mapeamento

Abstract

OpenStreetMap (OSM) is a collaborative mapping tool developed in the year 2004 by computer student Steve Coast of the University College London (UCL). Initially, the idea of the project was to map only the United Kingdom (UK) region, but soon aroused the interest of researchers from other countries. Since users actively participate in the processes of inclusion, change and exclusion of the data made available in the tool, it is necessary that there is a constant concern regarding the quality and consistency of them. Thus, the main objective of this work is to perform an analysis about the evolution of geographic data from Brazil inserted within this platform, based on the period between 2007 and 2016.

Keywords: OpenStreetMap, Geographic Data, Geographic Information Systems, Crowdsourcing, Mapping

Sumário

1	Introdução	1
1.1	Objetivo	2
1.1.1	Objetivos Específicos	2
1.2	Estrutura do Trabalho	2
2	Referencial Teórico	4
2.1	Bancos de Dados Espaciais	4
2.1.1	Modelagem de Dados Espaciais	5
2.1.2	O Modelo OMT-G	6
2.1.3	Representação Vetorial	8
2.1.4	Representação Matricial	9
2.1.5	Rede Triangular Irregular	10
2.1.6	O Consórcio Geoespacial Aberto	11
2.2	Sistemas de Informações Geográficas	11
2.2.1	Sistemas de Informações Geográficas Voluntárias	13
2.2.2	Sistemas de Referência de Coordenadas	15
2.3	<i>Crowdsourcing</i> - Contribuição Colaborativa	16
2.4	OpenStreetMap	17
2.4.1	Tipos de Objetos no OSM	18
2.4.2	O Fenômeno OpenStreetMap	19
2.4.3	Critérios de Qualidade no OSM	20
2.5	Trabalhos Relacionados	21
3	Desenvolvimento	23
3.1	Metodologia	23
3.1.1	Coleta dos Dados	23
3.1.2	Análise dos Dados	27
3.2	Arquitetura para Coleta e Visualização dos Dados	28

3.3 Ferramentas	29
3.3.1 PostgreSQL	29
3.3.2 PostGIS	30
3.3.3 Hstore	30
3.3.4 Osm2pgsql	31
3.3.5 OsmConvert	31
3.3.6 QGIS	32
4 Resultados	33
4.1 Dados Geográficos no Brasil	33
4.1.1 Atividades dos Usuários	33
4.1.2 Análises por Região e por Estado	37
4.1.3 Dados Geográficos e Concentração Populacional	41
4.1.4 Evolução de Estradas Brasileiras	42
4.2 Dados Geográficos no Distrito Federal	46
5 Conclusão e Trabalhos Futuros	52
Referências	53
Anexo	55
I Consultas SQL	56

Lista de Figuras

2.1	Exemplos de Classes do Tipo Geo-Objeto [6].	7
2.2	Diagrama de Classes do Modelo OMT-G [6].	8
2.3	Representação Vetorial do Polígono P [28].	9
2.4	Representação Matricial do Polígono P [28].	10
2.5	Representação de uma TIN [28].	10
2.6	Arquitetura de um Sistema de Informações Geográficas [21].	12
2.7	Camadas da Arquitetura de um SIG [18].	13
2.8	Esquema Representativo - <i>Crowdsourcing</i> [9].	17
2.9	Página Padrão do OpenStreetMap [15].	18
2.10	Usuários Registrados no OpenStreetMap.	20
3.1	Parte Inicial de um Arquivo no Formato <i>.osm</i>	25
3.2	<i>Menu</i> de Opções da Ferramenta Osmconvert.	25
3.3	Comando Executado na Ferramenta Osm2pgsql.	26
3.4	Estatísticas Geradas pela Ferramenta Osm2pgsql.	27
3.5	Arquitetura para Coleta e Visualização dos Dados.	29
4.1	Participação dos Colaboradores do OpenStreetMap por País.	34
4.2	Número de Usuários Ativos Mensais no OSM - Brasil a Cada Três Meses.	35
4.3	Usuários Colaboradores de Pontos, Linhas e Polígonos no OSM - Brasil.	36
4.4	Inclusão de Pontos, Linhas e Polígonos no OSM - Brasil.	37
4.5	Inserção de Pontos no OSM - Brasil, por Região.	38
4.6	Inserção de Pontos a Cada 1.000 Quilômetros Quadrados em Cada Região.	39
4.7	Inserção de Pontos a Cada 10.000 Habitantes em Cada Região.	39
4.8	Estados com a Maior Inserção de Pontos no OSM - Brasil até 2016.	40
4.9	Estados com a Menor Inserção de Pontos no OSM - Brasil até 2016.	41
4.10	Inserção de Pontos de Acordo com o Tamanho da População de Municípios.	42
4.11	Evolução de Estradas no OSM - Brasil por Categoria.	43
4.12	Evolução de Rodovias Federais e Estaduais no OSM - Brasil.	45
4.13	Nº de Usuários Colaboradores de Pontos, Linhas e Polígonos no OSM -DF.	46

4.14	Inserção de Pontos, Linhas e Polígonos no OSM - DF.	47
4.15	Tipos de Pontos Mais Inseridos no OSM - DF.	48
4.16	Evolução de Pontos e Polígonos do Tipo “Escola” na Região da Asa Norte.	49
4.17	Evolução de Estradas no OSM - DF por Categoria até 2016.	50
4.18	Evolução dos Dados do Tipo Linha no Distrito Federal.	51

Lista de Tabelas

3.1	Tamanho dos Arquivos Manipulados.	24
4.1	Distribuição de Pontos no OSM - Brasil, por Região.	38
4.2	Distribuição de Pontos Inseridos em Municípios, Conforme a População. . .	42
4.3	Distribuição de Estradas no OSM - Brasil por Categoria até 2016.	44
4.4	Distribuição de Pontos no OSM - DF por Categoria até 2016.	47
4.5	Distribuição de Estradas no OSM - DF por Categoria até 2016.	49

Capítulo 1

Introdução

Realizar o mapeamento da superfície terrestre faz parte da curiosidade humana, sendo dessa forma uma atividade bastante antiga. As sociedades pré-históricas já retratavam alguns elementos geográficos nas paredes das cavernas, e esse mapeamento primitivo servia não somente para localização e orientação dos indivíduos, mas também como forma de expansão das diferentes civilizações que ocuparam o planeta [25].

Ao longo do tempo, a forma de realizar o mapeamento terrestre tem se modificado de modo significativo, sendo que até há pouco tempo essa tarefa costumava ser restrita a alguns grupos específicos, como geógrafos e agências governamentais especializadas. Nos últimos anos, entretanto, surgiram algumas ferramentas que facilitaram essa atividade, permitindo que simples usuários pudessem inserir pontos e rotas de seus interesses a um mapa de forma computadorizada [12].

Em computação, mapas são armazenados em um tipo de banco de dados específico, denominado de Bancos de Dados Espaciais (ou Geográficos). Por sua vez, esses dados são manipulados através de um Sistema de Informação Geográfica (SIG). De modo particular, um Sistema de Informação Geográfica Voluntária (SIGV) utiliza as informações que seus usuários disponibilizam de forma colaborativa.

No contexto de utilizar ferramentas cada vez mais colaborativas e tendo como inspiração os mesmos princípios que deram origem à Wikipedia [30], a ferramenta OpenStreetMap (OSM) foi desenvolvida no ano de 2004, pelo estudante de computação Steve Coast, da *University College London (UCL)*. A ideia inicial do projeto era realizar o mapeamento apenas da região do Reino Unido, mas logo despertou o interesse de estudantes e pesquisadores de outros países [26].

Um dos grandes desafios em relação às ferramentas colaborativas é o fato de que elas necessitam da participação ativa dos usuários para a inclusão de novas informações à base de dados. Dessa forma, surgiu a Regra 90-9-1, responsável por descrever a chamada desigualdade de participação presente nesse tipo de ferramenta. Segundo essa regra, 90%

dos usuários apenas consomem o serviço, sem qualquer tipo de colaboração ativa; 9% colaboram de forma esporádica, e somente 1% correspondem aos membros que efetivamente colaboram com esse tipo de projeto [24].

Assim, a proposta desta monografia é elaborar uma análise sobre a evolução dos dados geográficos colaborativos no Brasil, entre o período de 2007 a 2016, observando-se como foi realizada a inserção desses dados dentro da plataforma do OSM. Neis *et al.*[23] desenvolveram um trabalho semelhante, porém apresentando como fonte os dados da Alemanha em um período de cinco anos.

1.1 Objetivo

O objetivo principal deste trabalho é realizar uma análise sobre a evolução dos dados geográficos no Brasil ao longo de um período de aproximadamente dez anos, observando-se como foi feita a inserção desses dados dentro da ferramenta OpenStreetMap (OSM).

1.1.1 Objetivos Específicos

Constituem objetivos específicos deste trabalho:

- Desenvolver uma arquitetura para a coleta e a visualização de dados a partir da ferramenta OpenStreetMap;
- Analisar a quantidade de usuários ativos presentes na ferramenta OSM no Brasil;
- Verificar a quantidade de dados presentes por região administrativa na ferramenta OSM no Brasil;
- Realizar análises em relação aos tipos de dados ponto, linha e polígono na ferramenta OpenStreetMap no Brasil;
- Realizar análises em relação aos dados geográficos presentes na ferramenta OSM correspondentes à região do Distrito Federal.

1.2 Estrutura do Trabalho

Os próximos capítulos da presente monografia estão dispostos da seguinte forma:

- Capítulo 2 - Fundamentação teórica em relação aos principais conceitos envolvidos para a elaboração deste trabalho, apresentando as definições de Bancos de Dados Espaciais, Sistemas de Informações Geográficas e *Crowdsourcing*. Além disso, o capítulo realiza uma introdução sobre a ferramenta OpenStreetMap.

- Capítulo 3 - Apresenta como este trabalho foi desenvolvido, demonstrando a metodologia utilizada, quais ferramentas foram necessárias e as etapas do desenvolvimento para a coleta e a análise dos dados.
- Capítulo 4 - Apresenta os principais resultados verificados em relação aos dados geográficos do Brasil, colocando-os na forma de gráficos, imagens e tabelas.
- Capítulo 5 - Destinado a apresentar as conclusões observadas e alguns trabalhos futuros.

Capítulo 2

Referencial Teórico

Este capítulo é dedicado a apresentar os principais conceitos envolvidos para a elaboração deste trabalho, tais como Bancos de Dados Espaciais, Sistemas de Informações Geográficas, *Crowdsourcing*, além de realizar uma introdução sobre a ferramenta OpenStreetMap. Além disso, o capítulo expõe alguns trabalhos já realizados com a temática envolvida.

2.1 Bancos de Dados Espaciais

Bancos de Dados Espaciais, também denominados de Bancos de Dados Geográficos, são bancos de dados que realizam a manipulação de objetos em um espaço multidimensional. Por exemplo, bancos de dados cartográficos representam uma instância de um banco de dados espaciais 2D e realizam o armazenamento de mapas, os quais por sua vez podem representar países, estados, cidades, rios, estradas, mares e assim por diante [22].

Por outro lado, os bancos de dados meteorológicos, para informações de clima, representam uma instância de um banco de dados espaciais 3D, pois as temperaturas e outras informações meteorológicas estão relacionadas a pontos do espaço tridimensional, uma vez que dependem não somente de informações acerca da latitude e longitude, mas também de informações acerca da altitude [22].

Para a representação de objetos em um Banco de Dados Geográficos, frequentemente é utilizado o Espaço Euclidiano, ou seja, um espaço vetorial real de dimensão finita. Isso significa que um ponto no plano será dado por um par de números reais. Por outro lado, os sistemas computacionais existentes trabalham com aproximações finitas e limitadas, o que dificulta o processo de modelagem desses tipos de bancos de dados e pode trazer alguns problemas na representação. Por exemplo, o ponto de intersecção entre duas linhas pode ser arredondado para o ponto mais próximo representável, não correspondendo

necessariamente às coordenadas geográficas do mundo real [13]. A seção a seguir busca detalhar um pouco mais de que forma é realizada a modelagem de dados geográficos.

2.1.1 Modelagem de Dados Espaciais

Dados espaciais necessitam de uma modelagem própria e, nesse sentido, existem dois diferentes aspectos que precisam ser representados em um Banco de Dados Geográficos [13]:

1. **Objetos no espaço:** Há o interesse de se representar entidades distintas dispostas no espaço, cada uma das quais com a sua própria descrição da respectiva geometria;
2. **Espaço:** É necessário que o espaço propriamente dito também seja inserido na representação, colocando-se em um mapa cada ponto desse espaço.

O primeiro aspecto permite, por exemplo, a modelagem de objetos como cidades, florestas ou rios. Já o segundo aspecto depende da temática que envolve o mapa representado, podendo corresponder à área de um país, distrito ou até mesmo ao mapa-mundi. Em um Banco de Dados Espaciais, esses dois aspectos são conciliados a fim de proporcionar a modelagem de objetos individuais, e coleções de objetos relacionados espacialmente entre si.

Um modelo de dados geográficos consiste em uma representação limitada da realidade, delimitada pela natureza finita dos dispositivos computacionais. Dessa forma, o modelo deve possuir um conjunto bem definido de objetos e operações válidas sobre esses objetos. Dados espaciais podem ser modelados de diferentes formas, e a seguir encontram-se descritos alguns dos principais modelos de dados geográficos [10].

- **Tesselações Regulares do Espaço (*raster*)** - O espaço é dividido em um conjunto finito de áreas pequenas e regulares. Esse é um método simples e útil para a discretização do espaço, que pode ser feita por amostragem regular, ou por aproximação pela média entre pontos próximos;
- **Spaghetti** - Linhas e polígonos são armazenados como sequências de coordenadas de pontos. O modelo Spaghetti é utilizado em pacotes de cartografia automatizada, onde as informações sobre os relacionamentos entre as entidades não são importantes;
- **Modelo de dados topológico** - Alguns tipos de relacionamentos espaciais são armazenados, sendo que a ênfase principal é dada aos relacionamentos de conectividade entre as linhas, e aos relacionamentos de adjacência entre polígonos. Esse é o modelo empregado na maioria dos Sistemas de Informações Geográficas atuais [19].

2.1.2 O Modelo OMT-G

O modelo denominado *Object Modeling Technique for Geographic Applications (OMT-G)* é um modelo orientado a objetos para aplicações geográficas, capaz de fornecer as primitivas básicas para a modelagem das diferentes geometrias e topologias que compõem uma base de dados espaciais [7].

Esse modelo teve inspiração no paradigma da Orientação a Objetos, surgindo a partir dos diagramas de classes da linguagem de modelagem de dados *Unified Modeling Language (UML)* [27]. A partir daí, primitivas geográficas foram introduzidas com o objetivo de aumentar suas capacidades semânticas. Os pontos fortes desse modelo são a alta expressividade gráfica e sua capacidade de representação, uma vez que as anotações textuais foram substituídas pela representação de relações de forma explícita, evidenciando as diferentes interações entre objetos espaciais ou convencionais.

O modelo OMT-G possui duas classes básicas: Classes Georreferenciadas e Classes Convencionais. A utilização desses dois tipos distintos de classes permite que diferentes aplicações compartilhem dados não espaciais, facilitando o desenvolvimento de aplicações integradas e a reutilização de dados [6].

Assim, tanto as Classes Georreferenciadas quanto as Classes Convencionais podem ser especializadas, utilizando-se para isso o conceito de herança, advindo da orientação a objetos. O modelo OMT-G formaliza a especialização das Classes Georreferenciadas em classes do tipo Geo-Campo e Geo-Objeto [6]. As classes do tipo Geo-Campo representam objetos distribuídos de forma contínua pelo espaço, correspondendo a grandezas como tipo de solo, topografia e teor de minerais. Já as classes do tipo Geo-Objeto representam objetos geográficos individualizáveis, que possuem identificação com elementos do mundo real, tais como rios e postes de iluminação. Classes do tipo Geo-Objeto podem conter apenas geometria ou podem conter geometria e topologia. A Figura 2.1 representa essas duas categorias de classes do tipo Geo-Objeto.

Dessa forma, pode-se dizer que um Geo-Objeto conterá apenas geometria no caso em que possuir somente propriedades acerca de suas características geométricas, denotando o tipo desse objeto na representação. Por outro lado, um Geo-objeto conterá geometria e topologia no caso em que possuir outras propriedades associadas, tais como o fluxo de direção desse objeto no espaço.

Os relacionamentos entre as classes de objetos do modelo OMT-G podem ser de três tipos: associação simples, relacionamento espacial ou relação de rede. Um relacionamento do tipo associação simples, por exemplo, expressa a relação firmada entre objetos convencionais ou entre um objeto convencional e um objeto espacial, sendo expresso por uma linha contínua entre as classes de objetos relacionadas [6] [8].

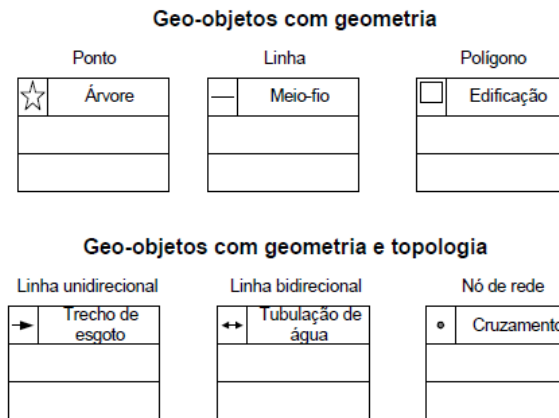


Figura 2.1: Exemplos de Classes do Tipo Geo-Objeto [6].

Um relacionamento do tipo espacial relaciona duas classes de objetos geográficos, sendo expresso por uma linha pontilhada simples entre as classes de objetos relacionadas. Por outro lado, uma relação de rede representa uma ligação lógica entre classes, em que os objetos se relacionam entre si através de nós ou arcos conectados e é representada por duas linhas pontilhadas paralelas entre as classes de objetos relacionadas [6] [8].

A Figura 2.2 representa um diagrama de classes do modelo OMT-G, no qual estão representados relacionamentos do tipo associação simples, relacionamento espacial e relacionamento de rede entre diferentes classes de objetos [6]. Nessa figura, percebe-se que entre a classe Linha de Ônibus e Ponto de Ônibus, o relacionamento é do tipo associação simples tendo em vista que relaciona uma Classe Convencional (Linha de Ônibus) a uma Classe Georreferenciada (Ponto de Ônibus). Já entre a Classe Ponto de Ônibus e as classes Endereço e Divisão Municipal, tem-se dois relacionamentos do tipo espacial, conectando Classes Georreferenciadas. As Classes Ponto de Ônibus e Área de Referência possuem um relacionamento de rede, denotando uma associação lógica entre essas classes.

Ainda na Figura 2.2, é possível visualizar a cardinalidade dos relacionamentos entre as classes de objetos, representando o número de instâncias de uma classe que podem estar associadas a uma instância da outra classe.

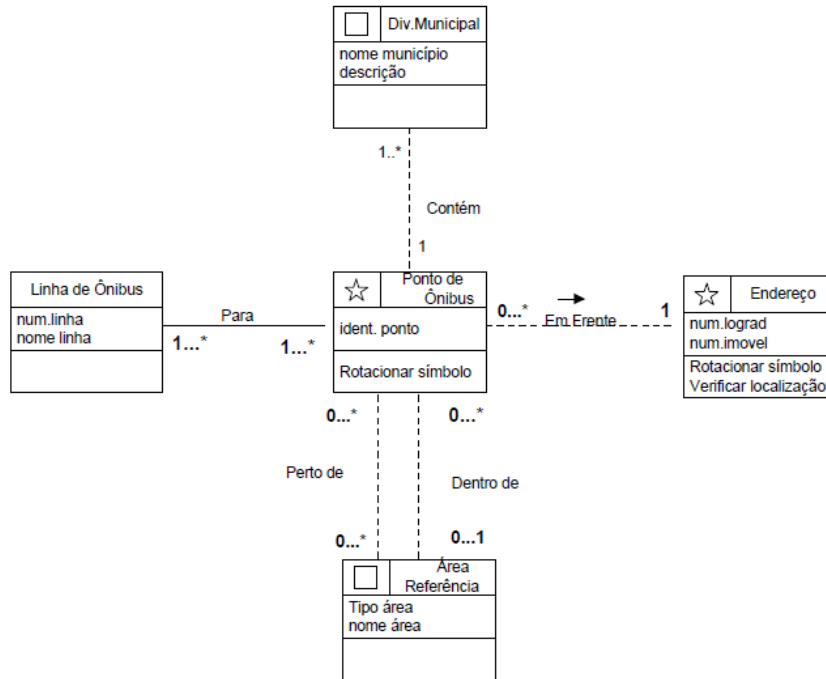


Figura 2.2: Diagrama de Classes do Modelo OMT-G [6].

2.1.3 Representação Vetorial

Os dados espaciais podem ser representados em diferentes formatos. Em geral, dados no formato vetorial são dados baseados em coordenadas geográficas, e isso quer dizer que cada objeto será representado por meio de um conjunto de coordenadas (x,y) ou (x,y,z) . Os objetos podem conter também outras informações adicionais por meio de seus diferentes atributos, armazenados no formato de *tags* [21].

Para a modelagem dos objetos individuais nesse tipo de representação, existe a noção de três abstrações fundamentais: o ponto, a linha e o polígono. Dessa forma, tem-se que:

- Um ponto simboliza um objeto cuja localização é relevante, mas cuja área é desconsiderada na representação;
- Uma linha (neste contexto sempre entendida como uma curva no espaço, geralmente representada por uma sequência de segmentos de linha) é a abstração básica para caminhos entre dois pontos. Geralmente, estradas, rios, cabos de telefonia e eletricidade são objetos espaciais modelados como linhas em um Banco de Dados Espaciais;
- Um polígono representa um objeto cuja área é relevante na representação, demarcando uma região bem definida. Essa região inclusive pode conter espaços vazios ou vários pedaços disjuntos.

Na representação vetorial, um polígono é definido por um conjunto de coordenadas geográficas, sendo que cada coordenada corresponderá a um vértice desse polígono em um espaço bidimensional ou tridimensional. Dessa forma, o polígono definido por $P = \langle [4,4],[6,1],[3,0],[0,2],[2,2] \rangle$ terá sua representação no espaço xy ilustrada pela Figura 2.3.

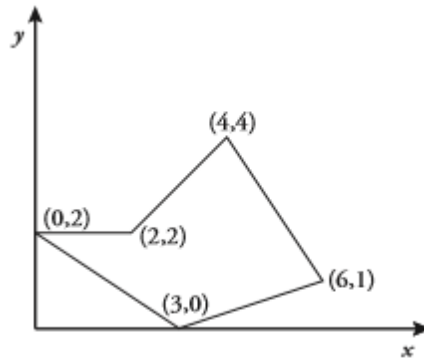


Figura 2.3: Representação Vetorial do Polígono P [28].

2.1.4 Representação Matricial

Dados no formato matricial (ou *raster*) são representados no formato de um *grid*, ou seja, um conjunto de linhas horizontais e verticais formando uma espécie de matriz. Geralmente, são dados provenientes de sensoriamento remoto ou imagens digitalizadas, produzidos por dispositivos eletrônicos ou satélites. Diferentemente dos dados vetoriais, os dados *raster* não contêm um registro em um banco de dados associado com cada célula, uma vez que os dados são geocodificados para cada *pixel* da resolução da imagem. Dessa maneira, é possível fazer a representação de paisagens com grandes variações de cores ou texturas, como ocorre em regiões de plantações, por exemplo [21].

Nesse caso, um polígono é definido não por um conjunto de coordenadas, mas sim por um conjunto de células do *grid*. O polígono definido por $P = \langle 5,12,13,14,17,18,19,20,21,22,26,27,28,29,30,31,35,36,37,38 \rangle$ terá a sua representação no espaço xy ilustrada pela Figura 2.4.

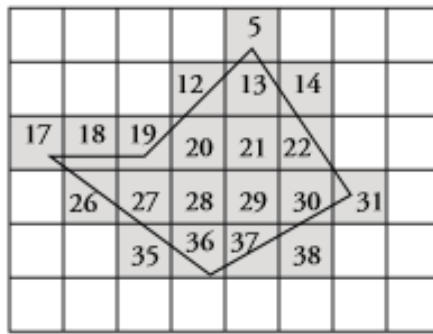


Figura 2.4: Representação Matricial do Polígono P [28].

2.1.5 Rede Triangular Irregular

Outra forma de representação de dados espaciais é através de uma Rede Triangular Irregular, do inglês *Triangulated Irregular Network (TIN)*¹. Nesse tipo de representação, o espaço bidimensional ou tridimensional é dividido em partições no formato triangular, conforme pode ser observado na Figura 2.5.

A principal motivação para o uso de uma TIN são as múltiplas resoluções oferecidas por um domínio irregular. Isso se traduz em economia computacional, uma vez que o número de nós pode ser significativamente reduzido. Outra vantagem é que as características lineares podem ser preservadas com precisão na malha, o que permite a imitação de linhas de ruptura do terreno, redes de fluxo e a demarcação de limites entre regiões heterogêneas [17].

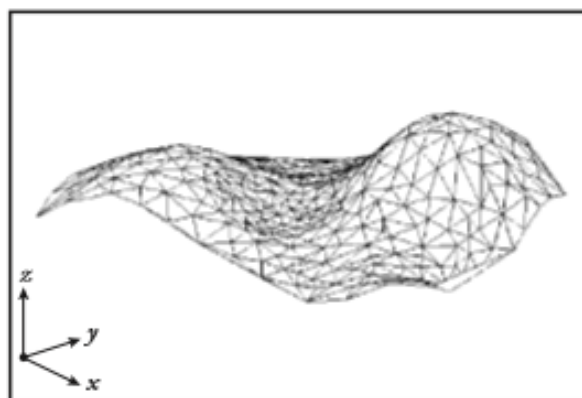


Figura 2.5: Representação de uma TIN [28].

¹https://en.wikipedia.org/wiki/Triangulated_irregular_network Acessado em julho de 2017.

2.1.6 O Consórcio Geoespacial Aberto

O Consórcio Geoespacial Aberto, do inglês *Open Geospatial Consortium (OGC)* é uma organização internacional que compreende mais de 300 organizações comerciais, governamentais, não-lucrativas e instituições de pesquisa do mundo inteiro. Algumas especificações mais importantes do OGC estão descritas a seguir [21]:

- **World Map Service (WMS)** – Auxilia no acesso a informações de mapas ou imagens georeferenciadas através da Internet. Os dados são colocados no formato de *grid* e acessados por um navegador *Web* ou Sistema de Informações Geográficas;
- **Web Feature Service (WFS)** – É similar ao formato de dados vetoriais e permite que o usuário obtenha características físico-geográficas a partir de um mapa, permitindo-se inclusive a edição desses dados;
- **Simple Feature for SQL (SFS)** – Auxilia no acesso a informações armazenadas no banco de dados. Na ferramenta QGIS, por exemplo, as camadas da extensão espacial POSTGIS são armazenadas em um banco de dados do Sistema Gerenciador de Bancos de Dados (SGBD) PostgreSQL. Além disso, o PostGIS permite a indexação de consultas e a filtragem dos dados em um banco de dados através da linguagem *Structured Query Language (SQL)*;
- **Geographic Mark-up Language (GML)** – Linguagem de marcação que pode ser utilizada para expressar as características geográficas. É um formato de intercâmbio aberto para transações geográficas na Internet.

2.2 Sistemas de Informações Geográficas

O termo Sistema de Informações Geográficas (SIG), ou *Geographic Information System (GIS)*, é aplicado aos sistemas que realizam o tratamento computacional de dados geográficos, armazenando geometrias e dados que são georeferenciados, ou seja, situados na superfície terrestre e representados em uma projeção cartográfica. Em geral, um SIG possui os seguintes componentes [13]:

- Interface com o usuário - Define como o sistema será operado e controlado;
- Entrada e integração de dados - Um SIG deve possuir mecanismos básicos de processamento de dados, o que envolve a parte de coleta e edição desses dados;
- Funções de processamento de gráficos e imagens - Uma vez que um SIG trabalha com elementos gráficos, é fundamental que seja capaz de manipular corretamente esses tipos de funções;

- Visualização e plotagem de dados em um mapa - A produção de mapas é uma atividade bastante recorrente em um SIG, uma vez que a representação visual dos dados é muito importante nesses tipos de sistemas;
- Armazenamento e recuperação dos dados por meio de um Sistema Gerenciador de Banco de Dados (SGBD) - Uma boa integração entre o SIG e um SGBD é fundamental a fim de facilitar o compartilhamento dos dados, recuperação através de mecanismos de *backup*, além de aumentar a disponibilidade e a integridade dos dados.

Os componentes de um Sistema de Informações Geográficas se relacionam entre si de acordo com uma arquitetura própria, conforme pode ser observado na Figura 2.6.

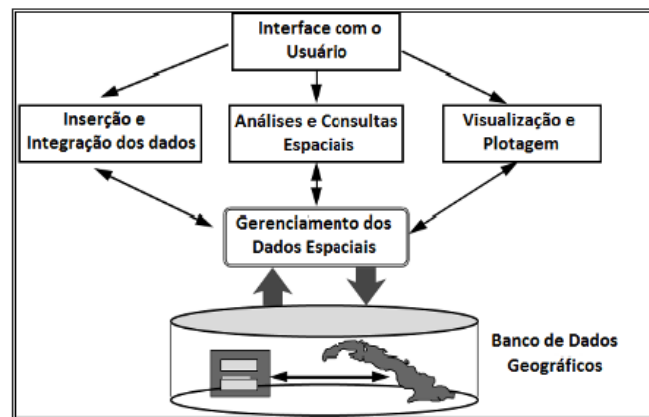


Figura 2.6: Arquitetura de um Sistema de Informações Geográficas [21].

A partir da Figura 2.6, observa-se que a arquitetura de um SIG é estruturada de uma forma hierárquica, apresentando três diferentes camadas:

1. Camada de Apresentação – Implementa a interface com o usuário do sistema, definindo como este será operado e controlado. Geralmente, os usuários acessam os dados de um SIG por meio de instruções específicas e os resultados podem ser visualizados através de uma ferramenta própria ou um módulo de visualização dentro do próprio SIG;
2. Camada de Aplicação – Provê a maior parte das funcionalidades de um SIG, incluindo as partes de inserção e integração dos dados, análises e consultas espaciais, além da visualização e plotagem dos dados na forma de mapas. Ela realiza a ligação entre as camadas de apresentação e de dados;

3. Camada de Dados – Realiza o gerenciamento dos dados espaciais através de um SGBD, o que envolve, portanto, operações de armazenamento e recuperação dos dados.

Portanto, a arquitetura de um SIG também pode ser descrita de forma mais simplificada por meio da Figura 2.7.

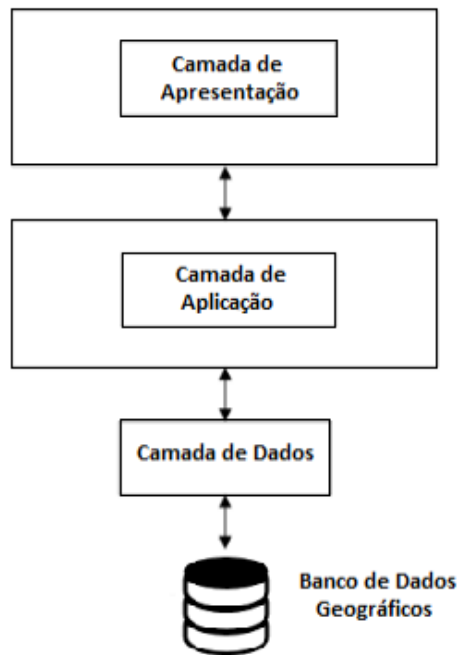


Figura 2.7: Camadas da Arquitetura de um SIG [18].

2.2.1 Sistemas de Informações Geográficas Voluntárias

Um Sistema de Informações Geográficas Voluntárias (SIGV) pode ser traduzido como um melhor aproveitamento das ferramentas utilizadas para a criação, a montagem e a disseminação de dados geográficos, fornecidos voluntariamente por um conjunto de indivíduos.

Na implementação de um SIGV, os indivíduos trabalham de forma independente, de acordo com as suas necessidades, e juntos formam uma organização própria de trabalho, conhecida como “*patchwork*”, segundo a qual uma espécie de miscelânea de informações vai sendo construída de forma contínua. Dessa forma, cada componente humano pode ser enxergado como um verdadeiro sensor geográfico, podendo atuar inclusive no alerta de possíveis emergências, tais como a ocorrência de terremotos ou furacões [12].

Para entender o funcionamento de um Sistema de Informações Geográficas Voluntárias, faz-se necessário destacar as principais ferramentas computacionais que tornaram esse tipo de tecnologia possível. Dessa maneira, pode-se citar as seguintes tecnologias [12]:

- **Web 2.0:** No começo dos anos 2000, novos protocolos de *Internet* foram desenvolvidos, os quais permitiram com que usuários pudessem ter acesso a bancos de dados armazenados em servidores *web*, podendo inclusive fazer a alteração de registros através de formulários. Tal fato permitiu a existência de *sites* cujo conteúdo é quase totalmente inserido por meio de seus usuários, com pouca moderação ou controle, e quase nenhuma restrição em relação à natureza do conteúdo inserido;
- **Georeferenciamento:** Para possibilitar a criação de dados geográficos pelo público em geral, é necessário dispor de um conjunto de ferramentas facilmente disponíveis a fim de identificar coordenadas geográficas (latitude e longitude) na superfície terrestre. Várias ferramentas utilizam-se de imagens capturadas por satélites, para assim criarem registros digitais de ruas inteiras e, em seguida, essas imagens são carregadas e compiladas na forma de mapas digitais;
- **Geotags (Etiquetas Geográficas):** Uma *geotag* ou etiqueta geográfica é um código normalizado que pode ser inserido em uma informação geográfica a fim de que possa ser melhor observada a sua localização. Muitas *geotags* foram utilizadas a fim de relacionar artigos da Wikipedia a objetos geográficos associados, por exemplo;
- **Global Positioning System (GPS):** O Sistema de Posicionamento Global é sem dúvida o primeiro sistema na história humana a permitir uma medição direta da posição de objetos na superfície terrestre. Os receptores GPS são fáceis de serem utilizados e fornecem uma estimativa instantânea da posição de um objeto, com acurácia muitas vezes melhor do que a precisão de 10 metros. Incorporado a sistemas de navegação, o GPS permite a localização de determinado veículo comparando-se ao conteúdo de um mapa digital. O GPS tem suscitado uma série de atividades envolvendo Sistemas de Informações Geográficas, destacando-se a criação de mapas digitais para pedestres, ciclistas ou veículos;
- **Recursos Gráficos:** Gráficos de alta qualidade constituem uma inovação relativamente recente na história da computação. A visualização dinâmica de objetos tridimensionais, como ocorre com o Google Earth² por exemplo, exige componentes de hardware com poderosos recursos gráficos, o que se tornou viável apenas no começo dos anos 2000;

²<https://earth.google.com/web/> Acessado em abril de 2017.

- **Comunicação Banda Larga:** O desenvolvimento dos SIGV não poderia se tornar possível sem um bom acesso à *Internet*, preferencialmente via uma conexão de alta capacidade. Atualmente, existe um grande número de pessoas com *Internet* via conexão banda larga, cuja infraestrutura pode fazer uso de satélites, cabos ou ainda aproveitar a estrutura utilizada para telefonia.

2.2.2 Sistemas de Referência de Coordenadas

Mapas tentam retratar a superfície terrestre, ou pelo menos parte dela, em um pedaço de papel plano ou tela de algum dispositivo eletrônico, no caso de um mapa virtual. Um Sistema de Referência de Coordenadas (SRC) então define, com a ajuda de coordenadas, como o mapa bidimensional projetado em um SIG estará relacionado a lugares reais na Terra. A decisão sobre qual projeção de mapa e sistema de referência de coordenadas deve ser utilizado depende da extensão regional da área em que se deseja trabalhar, da análise que deseja realizar e, muitas vezes, da disponibilidade dos dados [1]. Existem dois diferentes tipos de Sistemas de Referência de Coordenadas [5]:

- Os Sistemas de Coordenadas Globais ou Esféricas utilizam-se do padrão latitude-longitude. Estes são muitas vezes referidos como Sistemas de Coordenadas Geográficas.
- Já um Sistema de Coordenadas baseado em uma projeção cartesiana bidimensional (projeção de mapa), como o Universal Transversa de Mercator (UTM), projeção de Albers ou projeção de Robinson, fornecem os mecanismos básicos para fazer a projeção da superfície terrestre em um plano cartesiano bidimensional.

A ferramenta OpenStreetMap utiliza-se da projeção EPSG:3857, uma variante da projeção Mercator. Por esse motivo, houve a necessidade de atualizar o Identificador do Sistema de Referência Espacial, do inglês *Spatial Reference System Identifier (SRID)*, para esse sistema de coordenadas, especialmente para relacionar as tabelas com os dados do OSM com arquivos do tipo *shapefile*.

Um *shapefile* é um formato de armazenamento de dados vetoriais desenvolvido pela empresa americana *Environmental Systems Research Institute (Esri)*³. O formato é utilizado para armazenar a posição, forma e atributos de feições geográficas, sendo frequentemente utilizado para armazenar feições geográficas de tamanho extenso, com muitos dados associados.

³<http://www.esri.com/> Acessado em abril de 2017.

2.3 *Crowdsourcing* - Contribuição Colaborativa

O termo *crowdsourcing* pode ser definido como o ato de terceirizar tarefas originalmente realizadas dentro de uma organização, designando esse conjunto de tarefas a um grupo extenso e heterogêneo de usuários em potencial. O incentivo para a participação dos usuários, inclusive, pode ocorrer de forma monetária ou não monetária [16]. Em português, o termo pode ser descrito como *contribuição colaborativa* ou *colaboração coletiva*⁴.

Ferramentas de *crowdsourcing* contam com a contribuição de um grande número de indivíduos para a realização de tarefas específicas, e têm atraído cada vez mais atenção nos últimos anos. Muitas tarefas de *crowdsourcing* costumavam ser realizadas em *desktops* de forma *online* devido à sua conveniência e eficiência. No entanto, em algumas vezes este método mais tradicional pode não funcionar devido a requisitos especiais envolvendo localizações físicas reais de objetos específicos.

Dessa forma, um novo paradigma de coleta de dados, chamado de “*crowdsourcing* espacial” tem ganhado destaque nos últimos anos. O *crowdsourcing* espacial consiste em um conjunto de tarefas baseadas na localização física real do usuário e, dessa maneira, os usuários só conseguem completar determinada tarefa se estiverem próximos da localização física do objeto designado [31].

Muitos campos da Ciência da Computação têm utilizado a técnica de *crowdsourcing* como uma ferramenta de suporte, como é o caso das áreas de Mineração de Dados e Engenharia de *Software*. Porém, é na área de Bancos de Dados Espaciais onde esse tipo de ferramenta tem ganhado cada vez mais importância sobretudo na última década [9] [31] .

O desenvolvimento dos *smartphones* e da telefonia móvel contribuiu significativamente para o surgimento do *crowdsourcing* espacial, uma vez que permitiu com que usuários pudessem completar tarefas de acordo com a sua localização. Particularmente, muitos dispositivos móveis atuais utilizam-se de sensores embutidos para a coleta de informações, o que proporcionou o surgimento de um caso especial de *crowdsourcing* denominado de “*crowdsensing* espacial”.

Também deve-se saber que existe uma distinção entre os sistemas de *crowdsourcing* e as plataformas de *crowdsourcing*. Um sistema refere-se ao termo mais genérico, o que engloba uma parcela de ferramentas designadas para atender requisições de apenas um usuário por vez. Já as plataformas de *crowdsourcing* são sistemas que conseguem atender múltiplos usuários simultaneamente. A Figura 2.8 representa um esquema com os principais conceitos envolvidos com o termo *crowdsourcing* [9].

⁴<https://pt.wikipedia.org/wiki/Crowdsourcing> Acessado em julho de 2017.

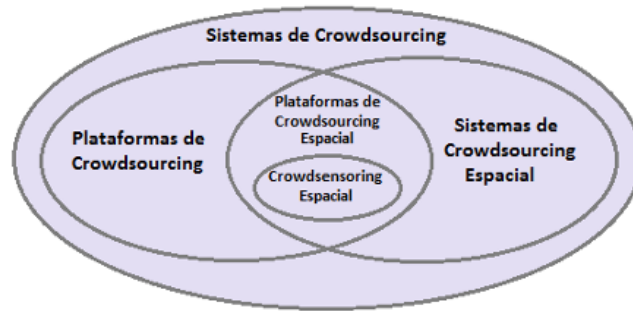


Figura 2.8: Esquema Representativo - *Crowdsourcing* [9].

2.4 OpenStreetMap

O OpenStreetMap (OSM) é um projeto de *crowdsourcing* espacial criado no ano de 2004 com o objetivo de se criar um mapa livre e editável do mundo construído por voluntários, em grande parte a partir do zero, e lançado com uma licença de conteúdo aberto [4].

Pode-se dizer que existem dois fatores primordiais que tornaram o OpenStreetMap uma ferramenta de sucesso: o primeiro foi a flexibilização das restrições em relação ao uso e acesso às informações geográficas do mundo inteiro; O segundo foi o barateamento dos dispositivos portáteis de navegação por satélite.

Além disso, os dados do OSM possuem uma grande vantagem por serem completamente livres, uma vez que estão dentro de uma licença de conteúdo aberto. Os dados na ferramenta são constantemente atualizados através de seus milhares de usuários cadastrados, os quais são capazes de inserir pontos relevantes ao mapa. O OSM possui ainda um significativo potencial para atrair voluntários do mundo todo, incluindo as regiões menos desenvolvidas do planeta, em que a obtenção de dados pode ser mais difícil para a maioria das empresas de mapeamento comercial [29].

A principal saída cartográfica do OpenStreetMap é apresentada na página web da ferramenta.⁵ A página utiliza-se da biblioteca AJAX⁶ para que o mapa possa ser atualizado em tempo real, e assim permitir uma interatividade maior com os usuários. Dessa forma, à medida que os usuários clicam em diferentes pontos do mapa, novos quadros são solicitados em segundo plano, sem precisar recarregar toda a página HTML. A Figura 2.9 representa a página padrão do OpenStreetMap, mostrando parte da área central de Londres, destacando-se em vermelho as principais ruas e avenidas situadas nessa parte da cidade.

⁵<https://www.openstreetmap.org/> Acessado em setembro de 2016.

⁶<http://ajaxian.com/by/topic/ajax> Acessado em julho de 2017.



Figura 2.9: Página Padrão do OpenStreetMap [15].

2.4.1 Tipos de Objetos no OSM

O OpenStreetMap trabalha com três tipos de objetos fundamentais: nós, caminhos e relações. Cada um desses objetos é descrito a seguir:

- **Nós:** Um nó representa um ponto na superfície terrestre, definido por sua latitude e longitude. Cada nó deve conter no mínimo um número de identificação (id) e um par de coordenadas. Nós podem ser utilizados para representar objetos como um banco de uma praça ou uma parada de ônibus, por exemplo. Os nós também podem ser utilizados como pontos de referência ao longo de caminhos, como para a representação de placas de trânsito ao longo de rodovias;
- **Caminhos:** Um caminho representa uma lista ordenada contendo de 2 a 2000 nós, definindo dessa forma o tipo “polilinha”. Caminhos representam estruturas lineares tais como rios e rodovias. Caminhos também são utilizados para definir limites de regiões fechadas como construções ou florestas e, nesse caso, o primeiro nó deve coincidir com o último nó. Áreas que contêm buracos, ou cuja fronteira supera 2000 nós, não podem ser representadas por um simples caminho, devendo ser representadas através de relações;
- **Relações:** Uma relação é uma estrutura de dados que documenta o relacionamento entre dois ou mais elementos (nós, caminhos ou outras relações). Alguns possíveis exemplos são:

- a. Relação de rota – Lista os caminhos que delimitam uma avenida, ciclovia ou rota de ônibus;
- b. Relação de restrição – Informa os locais em que o acesso de veículos, ciclistas ou pedestres é restrito.

O significado de muitas relações da ferramenta OSM é definido por meio de *tags* (etiquetas) específicas. Uma *tag* consiste em um par “chave-valor”, sendo que cada um desses campos contém uma *string Unicode* de até 255 caracteres.⁷

2.4.2 O Fenômeno OpenStreetMap

Em abril de 2006, a Fundação OpenStreetMap foi estabelecida com o objetivo de promover o crescimento, o desenvolvimento e a distribuição dos dados do OSM, permitindo o acesso e o compartilhamento desses dados entre os usuários do sistema.

Para a inserção dos dados dentro do OSM, muitos voluntários utilizam-se de uma aplicação escrita em Java, a qual é executada na própria página da ferramenta, ou então fazem uso de outros programas que utilizam dados gravados de GPS ou imagens aeroespaciais.⁸

Em dezembro de 2006, a empresa Yahoo⁹ permitiu que usuários do OpenStreetMap utilizassem suas fotografias tiradas por satélites para a produção de mapas dentro da ferramenta. Em novembro de 2010, o provedor de pesquisas Bing¹⁰ modificou a sua licença para também permitir o uso de suas imagens de satélites. Em 2012, como forma de recusa aos altos preços estipulados para o uso do Google Maps, a multinacional de tecnologia Apple lançou sua própria plataforma de mapeamento, utilizando dados do software proprietário TomTom, juntamente com dados do OSM.¹¹

O crescimento do OpenStreetMap aconteceu de forma bastante acelerada, atualmente ultrapassando a barreira dos três milhões de usuários registrados, conforme pode ser observado na Figura 2.10. No mês de janeiro de 2010, a ferramenta possuía aproximadamente 200.000 usuários registrados. Esse número foi multiplicado em cerca de dez vezes já no começo do ano de 2015, quando a ferramenta já contava com mais de dois milhões de usuários registrados¹².

⁷<http://wiki.openstreetmap.org/wiki/Elements> Acessado em setembro de 2016.

⁸http://wiki.openstreetmap.org/wiki/History_of_OpenStreetMap. Acessado em outubro de 2016.

⁹<https://br.yahoo.com/> Acessado em outubro de 2016.

¹⁰<https://www.bing.com/> Acessado em outubro de 2016

¹¹<https://en.wikipedia.org/wiki/OpenStreetMap>. Acessado em outubro de 2016.

¹²<http://wiki.openstreetmap.org/wiki/Stats> Acessado em fevereiro de 2017.

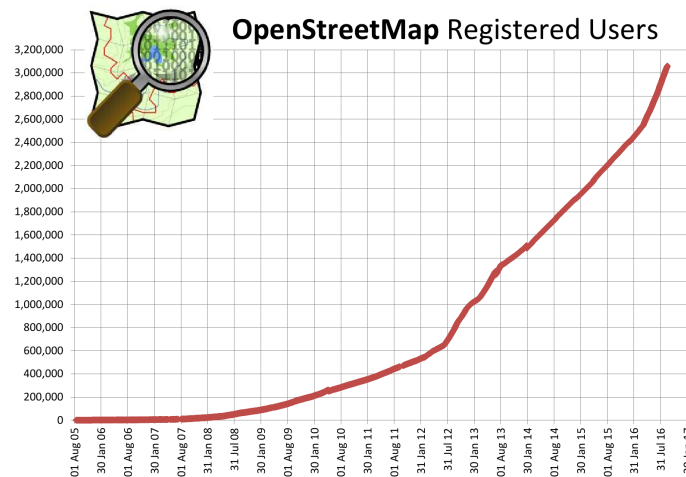


Figura 2.10: Usuários Registrados no OpenStreetMap.

2.4.3 Critérios de Qualidade no OSM

Por se tratar de uma ferramenta colaborativa, é necessário que haja uma constante preocupação com a qualidade dos dados inseridos no OpenStreetMap, a fim de verificar se os dados estão sendo inseridos de forma correta dentro da ferramenta. Nesse sentido, alguns dos principais critérios a serem analisados são os seguintes [29]:

- **Acurácia Posicional:** Representa a discrepância observada entre o ponto mapeado e o ponto de referência no mundo real. Para se verificar esse critério, muitos pesquisadores utilizam-se de algoritmos próprios para se determinar a porcentagem de objetos de uma base de dados que se encontram a uma certa distância de objetos que estão em outra base de dados com melhor acurácia;
- **Compleitude:** A completude é outro atributo muito importante para a análise de dados espaciais, uma vez que descreve a integralidade dos objetos e seus metadados. Para realizar a análise da completude de uma base de dados, geralmente, é criado um *grid* com uma resolução de 1 km, para que a partir daí seja realizada uma comparação dos dados que estão dentro de cada célula do *grid* com outra base de dados;
- **Acurácia Temporal:** Critério utilizado para descrever informações acerca da data de criação, modificação e exclusão de dados em uma determinada base de dados. A qualidade desse critério pode ser computada através do grau em que essas informações temporais descrevem ou não adequadamente as entidades espaciais. No OpenStreetMap, esse critério pode ser utilizado para comparar o histórico de determinado nó com o seu estado atual. Além disso, também é possível verificar a

existência de usuários mal-intencionados, capazes de inserir falsas informações propositalmente dentro do OSM;

- **Consistência Lógica:** Esse critério descreve o grau de confiabilidade das relações topológicas e lógicas entre os segmentos do conjunto de dados. Apesar de não possuir nenhum indicador quantitativo, esse critério é muito útil na visualização dos dados no OSM. Geralmente, o critério da Consistência Lógica está associado a dados do tipo Polígono, como por exemplo na definição de início, fim e junções de rodovias.

Além destes, outros critérios podem ser utilizados a fim de medir a qualidade dos dados no OpenStreetMap, dependendo do objetivo ao qual o pesquisador deseja alcançar.

2.5 Trabalhos Relacionados

Vários pesquisadores já propuseram investigações em relação à qualidade dos Sistemas de Informações Geográficas e, em particular, da ferramenta OpenStreetMap nos últimos anos. Um dos precursores dessas pesquisas foi o inglês Mordechai Haklay, o qual realizou um trabalho comparando a base de dados da ferramenta OSM com a base de dados de agências oficiais de Londres no ano de 2008 [14].

A partir daí, vários estudiosos continuaram nessa linha de pesquisa, e em 2009 os dados do OpenStreetMap na Alemanha foram comparados com a base de dados do software proprietário TomTom¹³. Outros trabalhos semelhantes foram realizados, porém comparando-se os dados do OSM com a base de dados de outros softwares. Um dos principais resultados obtidos nesses trabalhos foi a verificação de uma acentuada discrepância entre os dados de regiões urbanas e rurais, uma vez que foi verificada uma riqueza de detalhamento muito maior nas áreas urbanas comparando-se com os dados encontrados nas áreas rurais [34].

Uma abordagem parecida foi utilizada por pesquisadores franceses, os quais demonstraram as vantagens e a alta flexibilidade da ferramenta OpenStreetMap, apesar de também terem verificado uma grande heterogeneidade em relação aos dados coletados [11].

Em 2011, foram realizados os primeiros estudos acerca da qualidade dos dados do OpenStreetMap fora da Europa, quando pesquisadores americanos investigaram os dados presentes na ferramenta referentes ao estado da Flórida, realizando comparações com a base de dados dos softwares TomTom e NavTeq¹⁴. Como resultado, foi verificado algo contrário daquilo que foi analisado anteriormente nos estudos europeus sobre o OpenStreetMap, uma vez que na região da Flórida as áreas rurais possuíam um grau de detalhamento significativamente maior do que as áreas urbanas [32].

¹³http://www.tomtom.com/pt_br/ Acessado em maio de 2017.

¹⁴<https://here.com/en/navteq/> Acessado em maio de 2017.

Em 2012, Zielstra e Hochmair publicaram um estudo relacionando a ferramenta OpenStreetMap com o cálculo de menor rota para pedestres. Para tanto, os autores realizaram a análise de 1.000 pares aleatórios de pontos origem-destino em cidades da Alemanha e Estados Unidos, utilizando-se das bases de dados do OSM e dos softwares proprietários TomTom, NavTeq e ATKIS¹⁵. Os resultados dessa pesquisa mostraram que os dados do OSM apresentaram, em geral, rotas menores para pedestres, comparando-se um mesmo par origem-destino com a base de dados em ferramentas comerciais. Essa observação indica que a comunidade do OSM nesses dois países está ativa e de fato está disposta a colaborar com o desenvolvimento de uma rede abrangente de trilhas para pedestres [33].

Até a conclusão desta monografia, ainda faltavam estudos significativos em relação às informações contidas na ferramenta OpenStreetMap para a região correspondente ao Brasil. Dessa forma, este trabalho buscou uma metodologia semelhante a utilizada por Neis *et al*[23], a fim de se realizar uma análise em relação à evolução dos dados geográficos brasileiros em um período de aproximadamente dez anos. O próximo capítulo dedica-se a apresentar as etapas de desenvolvimento e metodologia utilizada para a realização deste trabalho, apresentando também a arquitetura proposta e quais ferramentas foram necessárias.

¹⁵<http://www.atkis.de/> Acessado em julho de 2017.

Capítulo 3

Desenvolvimento

Este capítulo dedica-se a apresentar como este trabalho foi desenvolvido, demonstrando a metodologia utilizada para a realização desta monografia, quais ferramentas foram necessárias e como foram elaboradas cada etapa do desenvolvimento deste trabalho.

3.1 Metodologia

A metodologia utilizada para a realização deste trabalho foi dividida em duas diferentes etapas. Na primeira etapa, foi realizada a parte da coleta dos dados a partir da ferramenta OpenStreetMap e inserção desses dados no Sistema Gerenciador de Bancos de Dados (SGBD) PostgreSQL. Já na segunda etapa, foi realizada a parte da análise dos dados, efetuada por meio de consultas na linguagem SQL, construções de gráficos e tabelas. As próximas seções dedicam-se a detalhar cada uma dessas etapas.

3.1.1 Coleta dos Dados

Para a coleta dos dados do OpenStreetMap, foi baixado o arquivo FullHistoryPlanet¹, contendo o histórico completo dos objetos armazenados na ferramenta OpenStreetMap de todos os países. Para nós, caminhos e relações criadas após a API 0.5 (Outubro de 2017)², o arquivo contém todas as versões desses objetos, mesmo que eles já tenham sido deletados. Entretanto, o arquivo não contém as versões anteriores a API 0.5 para nós e caminhos criados antes dessa API. O arquivo utiliza a mesma estrutura XML de um arquivo Planet³, porém além de conter as versões atuais de cada objeto, também contém suas versões anteriores, armazenando um histórico de cada objeto.

¹<https://planet.openstreetmap.org/planet/full-history/> Acessado em janeiro de 2017.

²http://wiki.openstreetmap.org/wiki/API_v0.5 Acessado em janeiro de 2017.

³<https://planet.osm.org/> Acessado em janeiro de 2017.

No *site* do projeto Planet OSM⁴, foi baixado o arquivo contendo os dados do OSM até o dia 23 de maio de 2017. O arquivo compactado com esses dados possuía um tamanho aproximado de 91 Gigabytes, porém ao descompactá-lo, foi gerado um arquivo que possuía algo em torno de 1,6 Terabyte, sendo necessário o auxílio de um HD externo com capacidade suficiente para armazená-lo. Após a extração dos dados referentes ao Brasil, por meio da utilização da ferramenta OsmConvert, foi gerado um novo arquivo com aproximadamente 20 GB, denominado de Brazil-History.osm. A Tabela 3.1 representa o tamanho dos arquivos manipulados a partir da realização deste trabalho.

Tabela 3.1: Tamanho dos Arquivos Manipulados.

Arquivo	Tamanho
FullHistoryPlanet (compactado)	91.645.475 KB $\approx 91GB$
FullHistoryPlanet (descompactado)	1.648.967.210 KB $\approx 1,6TB$
BrazilHistory.osm	20.135.420 KB $\approx 20,1GB$

O arquivo *FullHistoryPlanet* foi baixado no formato padrão *.osm*, pois era necessário capturar metadados importantes, como a informação a respeito da data em que os objetos foram inseridos à ferramenta (*tag osm_timestamp*) e o nome dos usuários responsáveis por fazer tais inserções (*tag osm_user*).

- A extensão *.osm*:

A extensão *.osm* utiliza-se da linguagem de marcação XML, apresentando os tipos primitivos do OpenStreetMap (nós, caminhos e relações) separados em diferentes blocos. Os dados do OSM estão originalmente na codificação UTF-8, tipo de codificação de caracteres em que é necessário apenas um *byte* para codificar os 128 caracteres ASCII. A *tag bounds* é utilizada para indicar as latitudes e as longitudes mínimas e máximas, demarcando a região em que os dados foram coletados. A Figura 3.1 representa a parte inicial de um arquivo *.osm*.

- Inserção dos dados no PostgreSQL:

Após a realização do *download* do arquivo *.osm*, a ferramenta Osmconvert foi utilizada a fim de realizar a delimitação da região referente ao Brasil, utilizando-se de um arquivo contendo o contorno do polígono desejado, nesse caso, o arquivo Brazil.poly. Arquivos *.poly* podem ser encontrados em diferentes *sites* na *Internet*. Neste trabalho foi utilizado o arquivo disponibilizado no *site* do projeto GeoFabrick⁵.

⁴<http://planet.osm.org/planet/full-history/> Acessado em maio de 2017

⁵<http://download.geofabrik.de/south-america/brazil.poly>. Acessado em janeiro de 2017.

```

<?xml version="1.0" encoding="UTF-8"?>
<osm version="0.6" generator="osmconvert 0.8.5" timestamp="2016-09-07T20:28:52Z">
  <bounds minlat="-35.46552" minlon="-74.09056" maxlat="5.522895" maxlon="-
27.67249"/>
  <node id="273316" lat="-26.9117879" lon="-48.6669852" version="18" timestamp="2016-
08-08T02:04:02Z" changeset="41314721" uid="2339960" user="adrianojbr">
    <tag k="is_in" v="Santa Catarina, Brazil"/>
    <tag k="is_in:continent" v="South America"/>
    <tag k="is_in:country" v="Brazil"/>
    <tag k="is_in:country_code" v="BR"/>
    <tag k="name" v="Itajai"/>
    <tag k="place" v="city"/>
    <tag k="population" v="205271"/>
    <tag k="rank" v="30"/>
    <tag k="ref:LOCODE" v="BRITJ"/>
  </node>

```

Figura 3.1: Parte Inicial de um Arquivo no Formato *.osm*.

A ferramenta Osmconvert pode ser utilizada em linha de comando, mas também permite com que o usuário navegue por um *menu* de opções, no qual o usuário poderá digitar o nome do arquivo a ser processado e, em seguida, poderá dentre outras opções converter o arquivo para um formato diferente, utilizar um arquivo *.poly* para delimitação de fronteiras e mostrar estatísticas sobre o arquivo.

A Figura 3.2 apresenta o *menu* de opções disponibilizado pela ferramenta Osmconvert. Entre as funções ofertadas pela ferramenta, destacam-se a conversão entre diferentes formatos de arquivos (*.pbx*, *.osm* e *.osm*) e a utilização de um arquivo de borda ou arquivo *.poly* para limitação da área geográfica, além da opção de disponibilizar estatísticas sobre um arquivo *.pbx*, *.osm* ou *.osm*.

```

What may I do with this file?

1  convert it to a different file format
2  use an OSM Changefile to update this file
3  use a border box to limit the geographical region
4  use a border polygon file to limit the geographical region
5  minimize file size by deleting author information
6  display statistics of the file
To options 3 or 4 you may also choose:
  a  keep ways complete, even if they cross the border
  b  keep ways and areas complete, even if they cross the border
Please enter the number of one or more functions you choose:

```

Figura 3.2: *Menu* de Opções da Ferramenta Osmconvert.

Para realizar a extração dos dados referentes ao Brasil, foi utilizada a opção número 4 (*use a border polygon file to limit the geographical region*), oferecida no *menu* de opções da Osmconvert. A partir daí, foi gerado o arquivo *BrazilHistory.osm*.

Feito isso, uma nova base de dados foi criada no PostgreSQL a fim de receber os dados coletados. As extensões PostGIS e Hstore foram incluídas nessa base de dados por meio

do comando “Create extension”. Em seguida, a ferramenta Osm2pgsql foi executada, por meio da aplicação da instrução ilustrada na Figura 3.3. Nesse comando, tem-se que:

- O argumento *extra-attributes* habilita *tags* especiais contidas nos dados do OpenStreetMap, tais como *osm_user*, *osm_uid*, *osm_version* e *osm_timestamp*;
- O argumento -d deve ser seguido do nome da base de dados no PostgreSQL, a qual receberá os dados coletados;
- O argumento -U deve ser seguido do nome do usuário da base de dados do PostgreSQL;
- O argumento -H deve ser seguido do *host* que se conectará à base de dados;
- O argumento -C deve ser seguido do tamanho da cache utilizada (em *Megabytes*);
- O argumento -W força o usuário a digitar a senha para realizar a conexão com a base de dados;
- O argumento *hstore* indica que os meta-dados do arquivo *.osm* deverão ser convertidos em *tags* através da extensão *hstore*;
- Por último, o usuário deverá indicar qual o local em que está contido o arquivo *.osm*, com os dados coletados da ferramenta OpenStreetMap.

```
osm2pgsql --extra-attributes -d BrasilHistorico -U postgres -H localhost -C 4096 -W --hstore C:\BrasilHistorico.osm
```

Figura 3.3: Comando Executado na Ferramenta Osm2pgsql.

A ferramenta Osm2pgsql realiza a conversão dos objetos típicos do OSM (nós, caminhos e relações) para os objetos geográficos típicos da maioria dos Sistemas de Informações Geográficas (pontos, linhas e polígonos), além de criar uma tabela específica para as rodovias (*roads*). A ferramenta gera, ao final da execução do comando, estatísticas como as apresentadas na Figura 3.4. Assim, na Figura 3.4, pode-se observar o número total de nós, caminhos e relações presentes no arquivo *.osm*.

```
Reading in file: C:\brazil2016.osm
Processing: Node(46734k 34.7k/s) Way(4443k 6.19k/s) Relation(92085 24.75/s) parse time: 5789s

Node stats: total(46734692), max(4390528868) in 1348s
Way stats: total(4443445), max(441290399) in 718s
Relation stats: total(92085), max(6568070) in 3720s
Committing transaction for planet_osm_point
Committing transaction for planet_osm_line
Committing transaction for planet_osm_polygon
Committing transaction for planet_osm_roads

Going over pending ways...
    1731893 ways are pending

Using 1 helper-processes
Helper process 0 out of 1 initialised
processing way (310k) at 0.23k/s
```

Figura 3.4: Estatísticas Geradas pela Ferramenta Osm2pgsql.

3.1.2 Análise dos Dados

A partir dos dados coletados, algumas consultas foram realizadas utilizando-se da linguagem SQL no SGBD PostgreSQL⁶. Já a visualização das consultas espaciais ficou a cargo da ferramenta QGIS⁷. Para as análises envolvendo regiões e estados brasileiros, assim como as análises envolvendo dados populacionais, foi necessário realizar o *download* de arquivos *shapefile*, para em seguida proceder ao relacionamento entre as tabelas com os dados do OpenStreetMap. As consultas realizadas foram as seguintes:

1. Número de usuários ativos mensais no OSM - Brasil a cada três meses;
2. Usuários colaboradores de pontos, linhas e polígonos no OSM - Brasil;
3. Inclusão de pontos, linhas e polígonos no OSM - Brasil;
4. Inserção de pontos no OSM - Brasil por região;
5. Inserção de pontos a cada 1.000 quilômetros quadrados em cada região brasileira;
6. Inserção de pontos a cada 10.000 habitantes em cada região brasileira;
7. Estados com a maior inserção de pontos no OSM - Brasil até 2016;
8. Estados com a menor inserção de pontos no OSM - Brasil até 2016;

⁶<https://www.postgresql.org/>. Acessado em outubro de 2016.

⁷<http://www.qgis.org/>. Acessado em outubro de 2016.

9. Inserção de pontos de acordo com o tamanho da população de municípios;
10. Evolução de estradas no OSM - Brasil por categoria;
11. Evolução de rodovias federais e estaduais no OSM - Brasil;
12. Número de usuários colaboradores de pontos, linhas e polígonos no OSM -DF;
13. Inserção de pontos, linhas e polígonos no OSM - DF;
14. Tipos de pontos mais inseridos no OSM;
15. Evolução de pontos e polígonos do tipo “Escola” na região da Asa Norte;
16. Evolução de estradas no OSM - DF por categoria até 2016;
17. Evolução dos dados do tipo linha no Distrito Federal.

3.2 Arquitetura para Coleta e Visualização dos Dados

A Figura 3.5 apresenta a arquitetura proposta para a coleta e visualização dos dados geográficos a partir da ferramenta OpenStreetMap. A arquitetura é composta de duas camadas principais, correspondentes às duas etapas do desenvolvimento deste trabalho. Dessa forma, tem-se uma camada para a coleta dos dados e uma camada para a visualização e análise dos dados.

Na camada referente à coleta dos dados, percebe-se a integração da ferramenta Osm2pgsql com o SGBD PostgreSQL e a extensão para dados espaciais PostGIS. Dessa forma, a ferramenta Osm2pgsql é capaz de realizar a conversão dos tipos primitivos contidos no arquivo .osm (nós, caminhos e relações) para os tipos primitivos básicos da maioria dos Sistemas de Informações Geográficas (pontos, linhas e polígonos).

Já no que se refere à parte de visualização dos dados, pode-se perceber a integração direta entre os dados geográficos contidos na extensão PostGIS e o software para visualização dos dados geográficos QGIS.

Nessa arquitetura, o SGBD PostgreSQL e a extensão PostGIS combinam um banco de dados objeto-relacional com uma estrutura espacial com suporte para objetos geográficos. O PostGIS utiliza recursos da especificação *Simple Features for SQL*, do Consórcio Geoespacial Aberto, e portanto, todos os dados são armazenados em formatos totalmente abertos. A ferramenta Osm2pgsql irá atuar na interconexão entre o PostgreSQL e a extensão PostGIS.

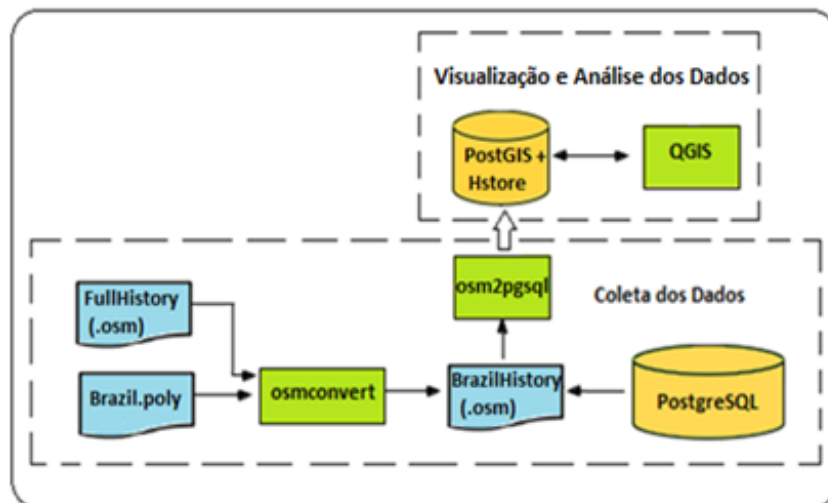


Figura 3.5: Arquitetura para Coleta e Visualização dos Dados.

3.3 Ferramentas

Para a realização deste trabalho foram utilizados os seguintes recursos computacionais: O SGBD PostgreSQL⁸, as extensões PostGIS⁹ e Hstore¹⁰, as ferramentas Osm2pgsql¹¹ e Osmconvert¹², além do software para visualização de dados geográficos QGIS¹³. Dessa forma, essas ferramentas encontram-se descritas nas próximas seções.

3.3.1 PostgreSQL

O PostgreSQL¹⁴ é um Sistema Gerenciador de Banco de Dados desenvolvido como projeto de código aberto, utilizado para bancos de dados relacionais. Criado há mais de quinze anos, atualmente o PostgreSQL detém uma forte reputação em relação aos critérios de confiabilidade, integridade dos dados e corretude. Além disso, trata-se de um SGBD que funciona em quase todos os sistemas operacionais existentes, incluindo Linux, Unix e Windows.

O PostgreSQL, originalmente chamado de Postgres, foi criado na *University of California at Berkeley* (UCB), pelo professor de Ciência da Computação Michael Stonebraker. O projeto do SGBD foi iniciado em 1985, como um trabalho futuro do seu predecessor Ingres, desenvolvido a partir de 1977.

⁸<https://www.postgresql.org/>. Acessado em outubro de 2016.

⁹<http://www.postgis.org/>. Acessado em outubro de 2016.

¹⁰<http://www.postgresqltutorial.com/postgresql-hstore/>. Acessado em outubro de 2016.

¹¹<http://wiki.openstreetmap.org/wiki/Osm2pgsql>. Acessado em outubro de 2016.

¹²<http://wiki.openstreetmap.org/wiki/Osmconvert>. Acessado em janeiro de 2017.

¹³<http://www.qgis.org/>. Acessado em outubro de 2016.

¹⁴<https://www.postgresql.org/>. Acessado em outubro de 2016.

O projeto inicial do PostgreSQL levou oito anos para ficar pronto, e nesse período foram introduzidas novas regras e procedimentos. A partir daí, novas funcionalidades foram sendo acrescentadas ao SGBD, permitindo uma integração cada vez melhor com a base de dados através da linguagem SQL.¹⁵

3.3.2 PostGIS

A extensão PostGIS adiciona suporte para objetos geográficos ao banco de dados objeto-relacional PostgreSQL. Na prática, essa extensão permite que o PostgreSQL seja usado como um banco de dados espacial para Sistemas de Informações Geográficas (SIG). A extensão PostGIS segue o padrão *Simple Features Specification for SQL*¹⁶, estabelecido pelo *Open Geospatial Consortium (OGC)*, estando atualmente em conformidade com esse padrão.

A criação da extensão PostGIS foi iniciada através do desenvolvimento da ferramenta *Refractions Research*, sendo desenvolvida como um projeto para bancos de dados espaciais *open source*. A extensão PostGIS continua a ser desenvolvida por um grupo de colaboradores liderados por um Comitê Diretor de Projetos e novos recursos continuam a ser adicionados frequentemente à ferramenta.

3.3.3 Hstore

A extensão Hstore¹⁷ implementa um novo tipo de dados ao SGBD PostgreSQL, denominado de *hstore*. O tipo de dados *hstore* armazena um par chave-valor em uma única célula de uma tabela do banco de dados. Esse tipo de dado é muito útil em diversos casos, podendo ser utilizado para o tratamento de dados semi-estruturados, ou linhas com muitos atributos que raramente são consultados. Na representação das *tags* do tipo chave-valor, tanto as chaves quanto os valores são implementados por meio de *strings* de caracteres.

Assim, neste projeto, a extensão *hstore* foi utilizada com o objetivo de realizar a conversão dos meta-dados contidos em um arquivo *.osm* para *tags* do tipo chave-valor. Para a realização deste trabalho, a extração da *tag* *osm_timestamp* era de suma importância a fim de que fossem devidamente realizadas as análises em relação à data em que os dados do OpenStreetMap foram inseridos dentro da ferramenta.

¹⁵<https://www.postgresql.org/about/history>. Acessado em outubro de 2016.

¹⁶<http://www.postgis.org/>. Acessado em outubro de 2016.

¹⁷<http://www.postgresqltutorial.com/postgresql-hstore/>. Acessado em outubro de 2016.

3.3.4 Osm2pgsql

A ferramenta Osm2pgsql¹⁸ é um programa baseado em linha de comando que converte os dados do OpenStreetMap para um banco de dados no PostgreSQL que esteja devidamente habilitado com a extensão PostGIS. Atualmente, a ferramenta não possui um único mantenedor, sendo mantida portanto por uma comunidade de colaboradores. Não há qualquer designação de versões “estáveis” ou “em desenvolvimento”, porém trata-se de uma ferramenta amplamente utilizada em ambientes de produção. A ferramenta Osm2pgsql encontra-se disponível para ambientes Linux, MAC OS X e Windows, sendo mais utilizada em sistemas Ubuntu ou Debian.

Alguns aspectos de como os dados do OSM serão convertidos para o PostgreSQL podem ser definidos a partir de um arquivo de estilo (*style file*). O arquivo de estilo padrão (*default style file*) acompanha a instalação da ferramenta Osm2pgsql e é compatível com a renderização na interface OSM Mapnik¹⁹. Com um arquivo de estilo, o usuário pode controlar como diferentes tipos de objetos e *tags* são mapeados para diferentes colunas e tipos de dados no banco de dados.

3.3.5 OsmConvert

A ferramenta OsmConvert²⁰ é um recurso muito útil para a conversão e o processamento de arquivos do OpenStreetMap. Ela pode ser utilizada tanto para realizar conversões entre os diferentes formatos em que os dados do OSM estão disponíveis (*.osm*, *.pbk*, *.osm*), como também para realizar o “recorte” de uma região específica, ao se informar as latitudes e as longitudes mínimas e máximas, ou através do fornecimento de um arquivo contendo o contorno do polígono desejado (arquivo *.poly*).

Ao extrair uma região geográfica (usando os argumentos *-b* ou *-B*), o arquivo de entrada deve conter os objetos ordenados por seu tipo: primeiro, todos os nós; em seguida, todos os caminhos, seguidos por todas as relações. Dentro de cada uma dessas seções, os objetos deverão estar ordenados por seu número de identificação (id) em ordem crescente. Os arquivos *.osm*, *.osm* e *.pbk*, geralmente, já se encontram adeptos a esta condição. Isso significa que provavelmente o usuário não precisará se preocupar com essa limitação. A ferramenta Osmconvert exibirá uma mensagem de erro, caso esta sequência esteja quebrada.²¹

¹⁸<http://wiki.openstreetmap.org/wiki/Osm2pgsql>. Acessado em outubro de 2016.

¹⁹<http://mapnik.org/> Acessado em maio de 2017

²⁰<http://wiki.openstreetmap.org/wiki/Osmconvert>. Acessado em janeiro de 2017.

²¹<http://manpages.ubuntu.com/manpages/zesty/man1/osmconvert.1.html> Acessado em maio de 2017.

3.3.6 QGIS

O software QGIS²² é um Sistema de Informações Geográficas gratuito e livre. O projeto nasceu em maio de 2002 e foi criado como um projeto na plataforma SourceForge²³ em junho do mesmo ano. O objetivo do projeto é fazer com que os SIGs (tradicionalmente softwares proprietários e caros) tornem a ser uma perspectiva viável para qualquer pessoa com acesso básico a um computador pessoal. A ferramenta QGIS, atualmente, pode ser executada na maioria das plataformas Unix, Windows e OS X. Desenvolvido na linguagem C++, o software tem uma interface gráfica amigável ao usuário.

Uma das vantagens do software QGIS é a sua distribuição sob a GNU *General Public License* (GPL). Isso significa que os desenvolvedores são capazes de inspecionar e modificar o código-fonte, ao mesmo tempo em que garantem que os usuários sempre terão acesso a um programa o qual é livre de custos e que pode ser modificado livremente²⁴.

O QGIS fornece um ótimo visualizador de dados geográficos, sendo utilizado por muitos pesquisadores e estudantes para as suas mais variadas necessidades de visualização desse tipo de dado. A ferramenta dá suporte à visualização de dados no formato matricial (*raster*) e no formato vetorial, podendo o usuário adicionar suporte a novos formatos facilmente utilizando-se de *plug-ins* dentro da ferramenta. Neste trabalho, o software QGIS permitiu a construção de figuras interessantes, tais como os mapas representados nas Figuras 4.12, 4.16 e 4.18, presentes no próximo capítulo desta monografia.

²²<http://www.qgis.org/>. Acessado em outubro de 2016.

²³<https://sourceforge.net/> Acessado em maio de 2017.

²⁴http://docs.qgis.org/2.14/en/docs/user_manual/preamble/foreword.html. Acessado em outubro de 2016.

Capítulo 4

Resultados

Este capítulo é dedicado à apresentação dos resultados obtidos através da elaboração deste trabalho, demonstrando a evolução dos dados geográficos do Brasil e em particular do Distrito Federal entre os anos de 2007 e 2016.

4.1 Dados Geográficos no Brasil

As próximas seções deste capítulo dedicam-se a apresentar os resultados obtidos referentes à evolução dos dados geográficos brasileiros em um período de aproximadamente dez anos. Primeiramente, foram realizadas análises em relação às atividades dos usuários da ferramenta OpenStreetMap, os quais atuaram na região correspondente ao Brasil durante o período observado. Em seguida, foram realizadas algumas análises em relação às regiões e aos estados brasileiros, análises em relação à concentração populacional de municípios e, por último, análises em relação à evolução das principais rodovias brasileiras.

4.1.1 Atividades dos Usuários

A comunidade brasileira no OpenStreetMap cresce a cada ano. Apesar da ferramenta ter sido criada no ano de 2004, os primeiros dados geográficos referentes ao Brasil foram acrescentados somente no ano de 2007. Em setembro de 2008, foi criado o primeiro tópico na Lista de Discussão do OSM no Brasil (Talk-BR)¹.

Em junho de 2013, foi criado o OpenStreetMap Fórum², contendo o sub-fórum “users: Brazil” de usuários brasileiros do OSM. Neste fórum os usuários podem acompanhar os tópicos simplesmente visitando o site, ou acessando o *feed* RSS apontado ao final de cada

¹<https://lists.openstreetmap.org/listinfo/talk-br> Acessado em novembro de 2016.

²<https://forum.openstreetmap.org/> Acessado em novembro de 2016.

página. O usuário autenticado no serviço pode ainda configurar o idioma da interface e eleger tópicos ou sub-fóruns para monitoramento por e-mail³.

Também com o intuito de permitir uma melhor conexão entre os diferentes usuários brasileiros da ferramenta OpenStreetMap, foram criados um chat IRC⁴, vários grupos no aplicativo de mensagens Telegram e uma sala de conferência virtual no Jabber-BR⁵.

A participação brasileira entre os colaboradores do OpenStreetMap, comparando-se com a participação de países europeus e dos Estados Unidos, pode parecer pequena: apenas 2%, como pode ser observado na Figura 4.1, extraída a partir do trabalho [3]. Porém, também a partir da Figura 4.1 observa-se que o Brasil já se encontra entre os países que mais colaboraram com a ferramenta OpenStreetMap nos últimos anos, igualando-se em participação a países como Espanha e Canadá.

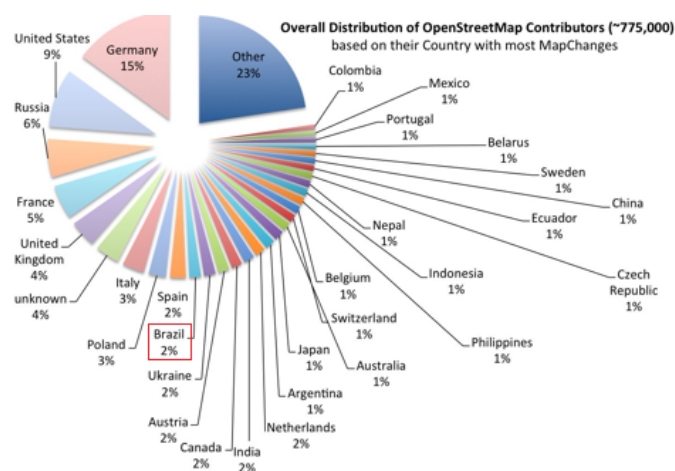


Figura 4.1: Participação dos Colaboradores do OpenStreetMap por País.

Os usuários do OpenStreetMap podem adicionar eventos ao calendário disponibilizado na Wiki do projeto⁶. Em 2013, por exemplo, aconteceram três eventos na cidade do Rio de Janeiro, uma conferência na cidade de Porto Alegre, um evento em Brasília e dois eventos na região do Recôncavo Baiano. Além disso, no mês de agosto foi promovida uma série de ações em virtude das comemorações ao nono aniversário da ferramenta OpenStreetMap⁷.

No ano de 2014, mais eventos aconteceram no Brasil. No mês de janeiro aconteceu a primeira *Mapping Party* na cidade de Salvador, em parceria com voluntários da Wikipédia⁸. Já no mês de maio, aconteceu uma maratona em preparação para a Copa do Mundo

³<http://wiki.openstreetmap.org/wiki/Pt:Contact> Acessado em novembro de 2016.

⁴<http://irc.openstreetmap.org/> Acessado em novembro de 2016.

⁵<xmpp:osm@conference.jabber-br.org> Acessado em novembro de 2016.

⁶https://wiki.openstreetmap.org/wiki/Main_Page Acessado em maio de 2017.

⁷http://wiki.openstreetmap.org/wiki/Past_Events_2013 Acessado em maio de 2017.

⁸http://wiki.openstreetmap.org/wiki/Mapping_Party_Salvador Acessado em maio de 2017.

de 2014, em que os usuários colaboradores concentraram seus esforços para realizarem o mapeamento próximo das áreas dos estádios onde ocorreriam jogos esportivos.⁹ Esse aumento no número de eventos do OSM nos anos de 2013 e 2014 pode ter contribuído em um aumento no número de usuários colaboradores ocorrido durante esse período.

Conforme pode ser observado em algumas estatísticas do OpenStreetMap, disponibilizadas na *Web*, o número de usuários ativos que colaboraram de alguma forma com a edição de dados geográficos do Brasil diariamente é algo bastante variável. Entre janeiro de 2012 e novembro de 2016, por exemplo, esse número variou de apenas 13 membros, ativos no dia 30 de março de 2013, até o pico de 93 membros ativos, alcançado no dia 19 de novembro de 2015 [2].

Apesar dessas oscilações, quando se realiza a análise em relação à atividade dos usuários de forma mensal, a tendência é majoritariamente crescente, conforme pode ser observado na Figura 4.2. Nessa figura está representado o número de usuários ativos mensais no OSM - Brasil, dividindo-se o período da análise em intervalos de três meses. Esse número chegou ao ápice em janeiro de 2014, apresentando uma queda até julho desse mesmo ano, quando voltou a crescer novamente.



Figura 4.2: Número de Usuários Ativos Mensais no OSM - Brasil a Cada Três Meses.

Na ferramenta OpenStreetMap os usuários podem colaborar por meio da inclusão, alteração e exclusão de pontos, linhas ou polígonos. Dessa forma, os usuários podem escolher qual o tipo de dado que desejam trabalhar, conforme sua necessidade ou disponibilidade. Assim, a Figura 4.3 apresenta o número de usuários colaboradores do OpenStreetMap no Brasil entre os anos de 2007 e 2016, agrupando-os pelo tipo de dado trabalhado. Também

⁹http://wiki.openstreetmap.org/wiki/Mapatona_Copa_do_Mundo_2014 Acessado em maio de 2017.

está descrito nessa figura o número total de usuários colaboradores no OSM - Brasil, o qual chega próximo da marca de 4.500 usuários registrados no ano de 2016.

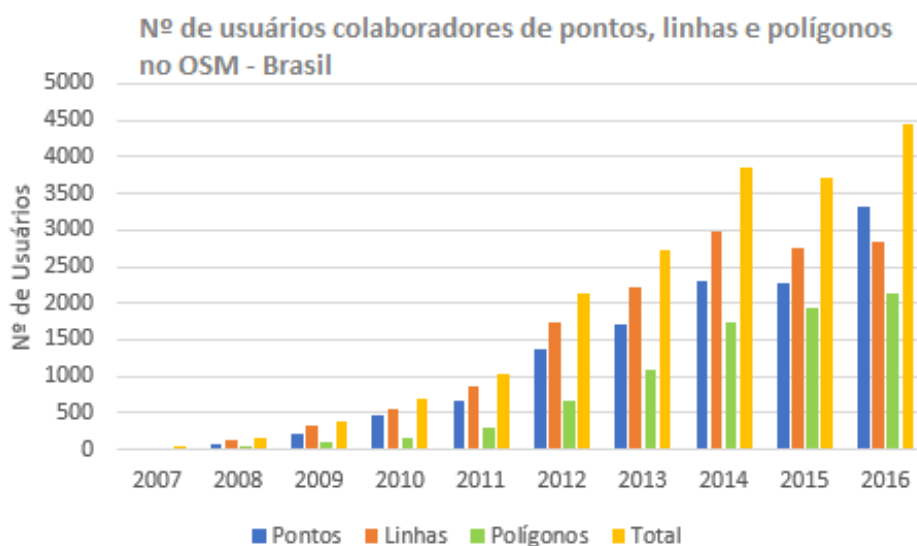


Figura 4.3: Usuários Colaboradores de Pontos, Linhas e Polígonos no OSM - Brasil.

Ainda na Figura 4.3, pode-se perceber que o número de usuários colaboradores de polígonos é o mais baixo ao comparar-se com o número de usuários colaboradores de pontos ou linhas. Tal fato pode ser atribuído à dificuldade que ainda existe para a inserção desse tipo de dado dentro da própria ferramenta OpenStreetMap. Para fazer a inclusão de um novo polígono na plataforma do OSM, o usuário deve traçar um caminho bem definido e fechado dentro de algum editor compatível com o OpenStreetMap, enquanto que para inserir um ponto é necessário realizar um simples clique.

A baixa participação de usuários colaboradores de polígonos refletia-se em uma baixa quantidade desse tipo de dado presente na ferramenta OpenStreetMap até o ano de 2015, descrita na Figura 4.4. Porém, esse quadro apresentou uma significativa melhora no ano de 2016, quando ocorreu um pico na inserção de polígonos no Brasil.

Um dado surpreendente é que aproximadamente 39% das inclusões de polígonos no ano de 2016 foram realizadas por um único usuário, denominado *smaprs_import*¹⁰, o qual participou de um projeto de mapeamento de edificações em Porto Alegre, e colaborou com o mapeamento na cidade do Rio de Janeiro em virtude dos Jogos Olímpicos de 2016.

De acordo com as Figuras 4.3 e 4.4, percebe-se que grande parte dos dados do OpenStreetMap correspondem ao tipo linha, e isso pode ser facilmente explicado uma vez que trata-se de uma ferramenta especializada na questão do roteamento, seja para o tráfego de veículos automotores, pedestres, transporte público ou ciclistas. A alta concentração

¹⁰http://www.openstreetmap.org/user/smaprs_import Acessado em junho de 2017.

de dados do tipo linha no OSM - Brasil torna-se ainda mais evidente após o salto na inclusão de dados ocorrido entre os anos de 2011 e 2012. Após 2012, a inclusão de linhas estabilizou-se na faixa entre 500 mil e 600 mil novas inclusões por ano.

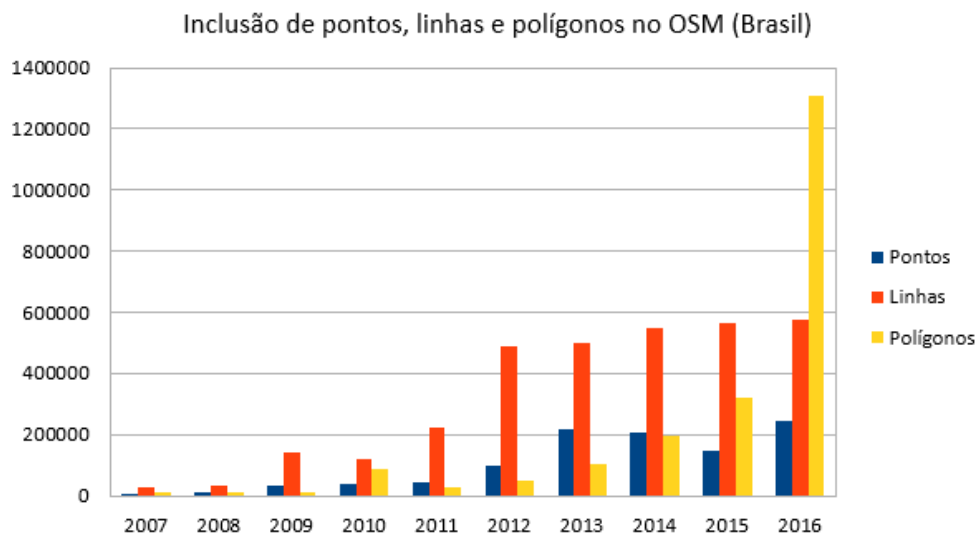


Figura 4.4: Inclusão de Pontos, Linhas e Polígonos no OSM - Brasil.

4.1.2 Análises por Região e por Estado

Em termos absolutos, os dados geográficos brasileiros foram incluídos na ferramenta OpenStreetMap de forma bastante heterogênea. A Figura 4.5 ilustra de que forma os objetos do tipo ponto foram inseridos no OSM - Brasil por região administrativa até o ano de 2016. Nessa figura destaca-se o grande volume de dados inseridos no ano de 2013 nas regiões Sul e Sudeste e, em menor escala, no ano de 2014 na região Nordeste. A região Norte é a que concentra a menor quantidade de pontos em quase todo o período observado.

Também na Figura 4.5 percebe-se que os dados das regiões Sul e Sudeste apresentaram tendências bem parecidas, tanto nos anos de crescimento como nos anos de decréscimo. Na região Centro-Oeste houve um período de ascensão entre os anos de 2012 e 2014, mas após o ano de 2014 a região apresentou uma tendência de queda, tendo inclusive registrado inserção menor de pontos do que a região Norte no ano de 2016.



Figura 4.5: Inserção de Pontos no OSM - Brasil, por Região.

A Tabela 4.1 mostra a distribuição de objetos do tipo ponto no OSM - Brasil, agrupados por região até o ano de 2016. A região Sudeste é a que teve maior participação na inclusão de pontos, com um percentual aproximado de 40%. A região Sul ficou em segundo lugar, impulsionada principalmente pela participação do estado do Rio Grande do Sul, o 2º colocado entre os estados em que foram incluídos mais pontos no OSM - Brasil até o ano de 2016, atrás apenas do estado de São Paulo. Também por meio da Tabela 4.1, pode-se confirmar a baixa inclusão de pontos na região Norte do país - apenas 6%.

Tabela 4.1: Distribuição de Pontos no OSM - Brasil, por Região.

Região	Participação
Sudeste	40%
Sul	28%
Nordeste	18%
Centro-Oeste	8%
Norte	6%

Quando se considera características peculiares de cada região brasileira, tais como a extensão territorial e o tamanho populacional, a análise aproxima-se mais da realidade de cada local. Dessa maneira, a Figura 4.6 ilustra a quantidade de pontos inseridos conforme a área de cada região. A Figura 4.6 é um pouco semelhante ao gráfico da Figura 4.5, porém pode-se perceber que o número de pontos inseridos por quilômetro quadrado na região Sul foi maior do que o número de pontos inseridos por quilômetro quadrado na região Sudeste no ano de 2013 (ano de pico de inserções de pontos no OSM - Brasil). Além disso,

percebe-se que a inserção de pontos por quilômetro quadrado apresentou uma tendência decrescente em todas as regiões brasileiras a partir do ano de 2014.



Figura 4.6: Inserção de Pontos a Cada 1.000 Quilômetros Quadrados em Cada Região.

Por outro lado, quando analisa-se a inserção de pontos no OSM - Brasil levando-se em conta a população de cada região, pode-se perceber que a imagem fica um pouco diferente. A partir da Figura 4.7, verifica-se que na região Centro-Oeste ocorreu uma maior inserção de pontos por habitante do que na região Sudeste entre 2012 e 2014. Além disso, também na Figura 4.7 observa-se a grande participação da região Sul para a inserção de pontos na ferramenta OpenStreetMap, sobretudo no ano de 2013.



Figura 4.7: Inserção de Pontos a Cada 10.000 Habitantes em Cada Região.

A partir da análise das Figuras 4.5 a 4.7, percebe-se que nas duas primeiras figuras há um distanciamento maior por parte da concentração de dados das regiões Sul e Sudeste, e as demais regiões. Por outro lado, quando se analisa o gráfico da Figura 4.7, percebe-se que a inclusão de pontos no OpenStreetMap ocorreu de forma mais ou menos uniforme, com exceção dos picos observados nos anos de 2013, 2014 e 2016. Em termos absolutos, verifica-se uma grande quantidade de pontos inseridos na região Nordeste. Porém, considerando-se a população, observa-se que a inclusão de pontos nessa região ocorreu de forma bastante semelhante à inserção de pontos ocorrida na região Norte do país.

Ao analisar-se a evolução de dados em relação aos cinco estados onde ocorreram as maiores inserções de pontos no OSM - Brasil (São Paulo, Rio Grande do Sul, Rio de Janeiro, Minas Gerais e Paraná), percebe-se que ocorreu um grande salto na inclusão de pontos para o estado de São Paulo nos anos de 2012 e 2013. Também no ano de 2013 ocorreu uma grande inclusão de pontos no estado do Rio Grande do Sul, superando até mesmo a quantidade de pontos inseridos no estado de São Paulo durante esse período. Tal fato está representado na Figura 4.8.

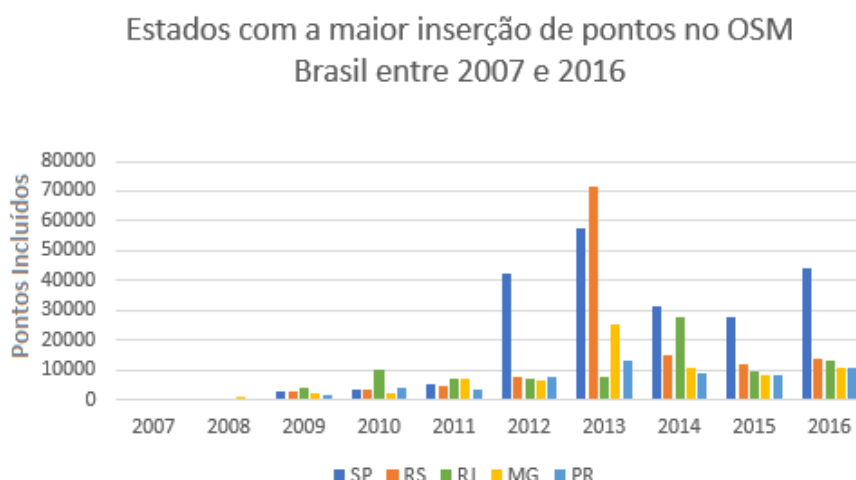


Figura 4.8: Estados com a Maior Inserção de Pontos no OSM - Brasil até 2016.

Todavia, quando se analisa a evolução entre os cinco estados onde ocorreram as menores inserções de pontos no OSM - Brasil até 2016 (Acre, Roraima, Rondônia, Amapá e Piauí), percebe-se que quatro desses estados são pertencentes à região Norte, com a exceção do estado do Piauí. Mesmo entre esses estados percebe-se alguns picos com uma inclusão maior de dados, destacando-se o pico alcançado de quase 3.500 pontos inseridos em 2014 no estado do Piauí. Também percebe-se uma melhora na inclusão de pontos para os estados da região Norte no ano de 2015, conforme pode ser observado na Figura 4.9.

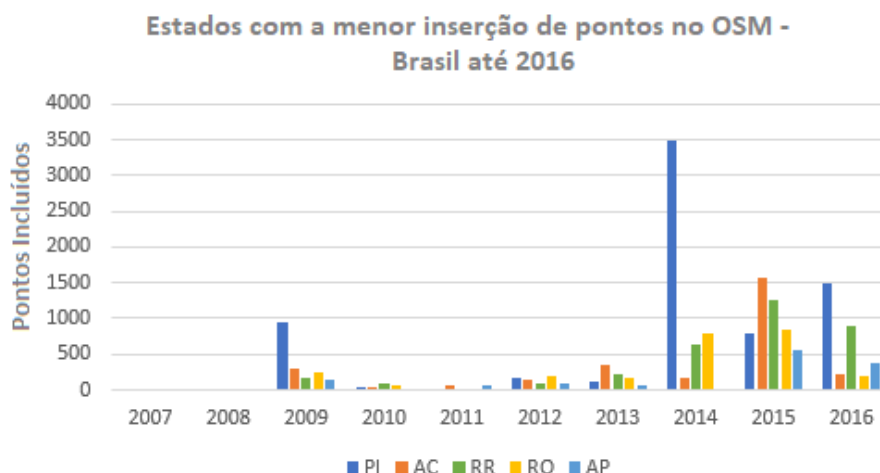


Figura 4.9: Estados com a Menor Inserção de Pontos no OSM - Brasil até 2016.

4.1.3 Dados Geográficos e Concentração Populacional

Intuitivamente, é fácil pensar que os dados geográficos inseridos em uma ferramenta colaborativa tenderiam a se concentrar mais nas áreas próximas aos grandes centros urbanos. Para verificar se tal fato realmente acontecia com o OpenStreetMap - Brasil, algumas consultas foram realizadas, dividindo-se os municípios brasileiros em diferentes grupos, conforme o tamanho de sua população. Dessa maneira, os municípios brasileiros foram divididos de acordo com o seguinte critério: o Grupo 1 corresponde aos municípios que possuíam até 10.000 habitantes segundo o Censo do IBGE de 2010; O Grupo 2 corresponde aos municípios que possuíam mais de 10.000 e menos de 50.000 habitantes; O Grupo 3 corresponde aos municípios que possuíam mais de 50.000 e menos de 100.000 habitantes; O Grupo 4 corresponde aos municípios que possuíam mais de 100.000 e menos de 500.000 habitantes; O Grupo 5 corresponde aos municípios que possuíam mais de 500.000 e menos de 1 milhão de habitantes; Por último, o Grupo 6 corresponde aos municípios que possuíam mais de 1 milhão de habitantes. A Tabela 4.2 ilustra a distribuição de pontos inseridos no OSM - Brasil, de acordo com o tamanho da população de municípios até o ano de 2016.

Conforme pode ser observado na Tabela 4.2, ocorreram maiores inserções de pontos nos municípios dos Grupos 2, 4 e 6. De forma surpreendente, o Grupo 2, pertencente aos municípios cuja população era de mais de 10.000 e menos de 50.000 habitantes, obteve uma grande participação na inserção de pontos no OSM - Brasil, com destaque para a participação de pequenas cidades do interior dos estados de São Paulo, Minas Gerais e Paraná.

Tabela 4.2: Distribuição de Pontos Inseridos em Municípios, Conforme a População.

Grupo	Participação
Grupo 1	9,31%
Grupo 2	23,33%
Grupo 3	8,41%
Grupo 4	24,55%
Grupo 5	7,95%
Grupo 6	26,45%

A Figura 4.10 ilustra a quantidade de pontos inseridos de acordo com o tamanho da população de municípios brasileiros. Assim, na Figura 4.10 destaca-se o grande salto no número de inclusões de pontos no ano de 2013 entre os municípios compreendidos pelo Grupo 6 (cidades com mais de 1 milhão de habitantes, correspondendo às grande metrópoles brasileiras), quando foi ultrapassada a barreira dos 90.000 pontos inseridos. Ainda nesta figura percebe-se a grande participação de grupos de cidades menores, destacando-se a participação dos Grupos 2 e 4.

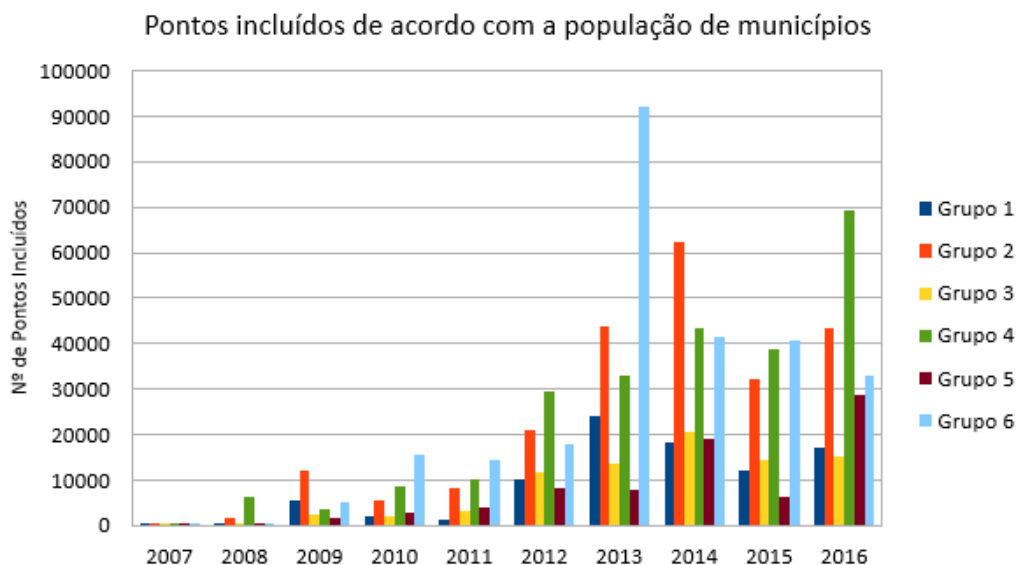


Figura 4.10: Inserção de Pontos de Acordo com o Tamanho da População de Municípios.

4.1.4 Evolução de Estradas Brasileiras

Conforme já foi falado, um tipo de dado bastante presente na ferramenta OpenStreet-Map no Brasil corresponde ao tipo linha, uma vez que existe uma ênfase maior na questão do roteamento. A partir desse fato, tornou-se interessante realizar algumas análises acerca desse tipo de dado, especialmente no que se refere à evolução das principais estradas bra-

sileiras entre os anos de 2007 e 2016. Dessa forma, a Figura 4.11 representa a injeção de estradas no OSM – Brasil em termos de quilômetros inseridos por categoria.

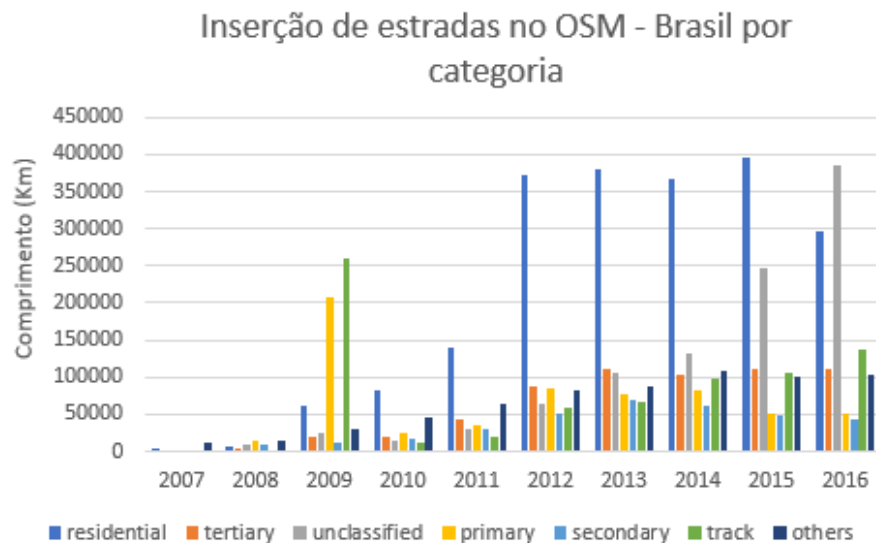


Figura 4.11: Evolução de Estradas no OSM - Brasil por Categoria.

A partir da Figura 4.11, percebe-se que no ano de 2009, período em que os dados ainda estavam sendo consolidados no OSM - Brasil, aconteceu um pico na inserção de estradas das categorias *primary* e *track*. A primeira categoria refere-se a rodovias primárias, tipicamente rodovias federais, no Brasil definidas pela sigla BR seguida por três algarismos (Ex.: BR-101). Já a categoria *track* refere-se a estradas de zonas rurais, geralmente sem pavimentação e utilizadas para fins agrícolas.

A partir de 2010, tem-se uma predominância na inserção de estradas próximas às áreas residenciais, e esse fato tornou-se mais evidente após o grande salto ocorrido entre os anos de 2011 e 2012, quando a inserção de estradas residenciais passou de menos de 150 mil quilômetros em 2011 para mais de 350 mil quilômetros de estradas inseridos no ano de 2012. Outro fato a ser considerado foi o aumento verificado na quantidade de estradas do tipo sem classificação (*unclassified*) a partir do ano de 2015, categoria tipicamente utilizada para estradas de tráfego local e para estradas utilizadas para ligar pequenas cidades ou vilarejos.

A Tabela 4.3 apresenta a distribuição de estradas por categoria no OSM - Brasil até o ano de 2015. Conforme pode ser observado, há uma predominância de rodovias residenciais (34,46%) e, em seguida, predomina-se rodovias dos tipos *unclassified* (16,58%) e *track* (12,35%).

Tabela 4.3: Distribuição de Estradas no OSM - Brasil por Categoria até 2016.

Categoria	Participação
Residencial	34,46%
Sem classificação	16,58%
Estrada rural	12,35%
Primária	10,34%
Secundária	5,59%
Terciária	10,06%
Outras	10,62%

Os dados do tipo linha estão organizados em diferentes categorias dentro da ferramenta OpenStreetMap. Para destacar-se apenas as rodovias brasileiras mais importantes, foi utilizada a chave *highway* (rodovia), quando seu valor correspondia a qualquer um desses valores: *motorway* (auto-estrada), *motorway_link* (ligação de auto-estrada), *trunk* (rodovia importante), *trunk_link* (ligação de rodovia importante), *primary* (rodovia primária) e *primary_link* (ligação de rodovia primária).

A partir da análise da Figura 4.12, pode-se observar a grande diferença existente em relação à quantidade de dados de rodovias brasileiras entre os anos 2008 e 2010, um intervalo de apenas dois anos. Também percebe-se uma heterogeneidade significativa entre os dados da parte leste e da parte oeste do país, visto que a parte mais oriental possui uma maior concentração de dados. Esse fato pode ser justificado pois é justamente nessa região onde estão situados os maiores centros urbanos do Brasil.

Conforme o passar dos anos, os dados nas regiões Sul, Sudeste, Centro-Oeste e Nordeste foram sendo acrescentados na ferramenta, com uma interligação cada vez maior por parte das rodovias entre si. Porém, até o final de 2015 havia poucos dados de rodovias na região Norte do país. Como uma possível explicação para isso, deve-se levar em consideração que a região Norte é a menos populosa do país, com pouco mais de 4 habitantes por quilômetro quadrado, conforme o Censo Demográfico de 2010 realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Além disso, a presença da Floresta Amazônica dificulta o acesso para o mapeamento da região.

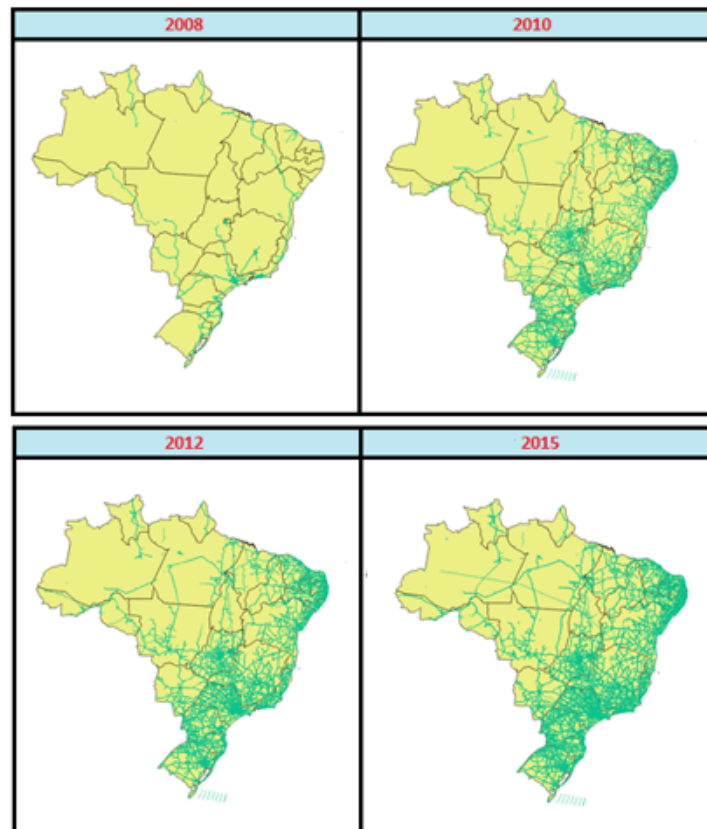


Figura 4.12: Evolução de Rodovias Federais e Estaduais no OSM - Brasil.

A próxima seção deste capítulo é destinada a realizar análises em relação à região do Distrito Federal (DF), unidade da federação em que está localizada a Universidade de Brasília (UnB). A metodologia e as consultas realizadas para a análise dos dados geográficos do DF foram semelhantes àsquelas utilizadas para as análises realizadas em nível nacional.

4.2 Dados Geográficos no Distrito Federal

Esta seção é dedicada a apresentar a evolução dos dados geográficos referentes ao Distrito Federal (DF) entre os anos de 2007 e 2016. Com a realização deste trabalho, percebe-se que o número de usuários que atuam no DF tem crescido de forma significativa, alcançando 239 usuários ativos no ano de 2016. A Figura 4.13 apresenta a evolução de usuários ativos na ferramenta OpenStreetMap, realizando-se a separação pelo tipo de dado editado. Dessa forma, obtém-se o número de usuários colaboradores de pontos, linhas e polígonos que atuaram na região correspondente ao DF em um período de dez anos.

Conforme pode ser observado na Figura 4.13, existem muitas semelhanças entre a atividade dos usuários que atuaram no DF e a tendência nacional. Assim como no nível nacional, a maior parte dos usuários do DF colaboraram com a edição de linhas, enquanto poucos usuários trabalharam com a edição de polígonos na ferramenta. Entretanto, no Distrito Federal ocorreram mais oscilações no número de usuários colaboradores, havendo alternância entre anos de queda e anos de ascensão, conforme também pode ser observado na Figura 4.13.

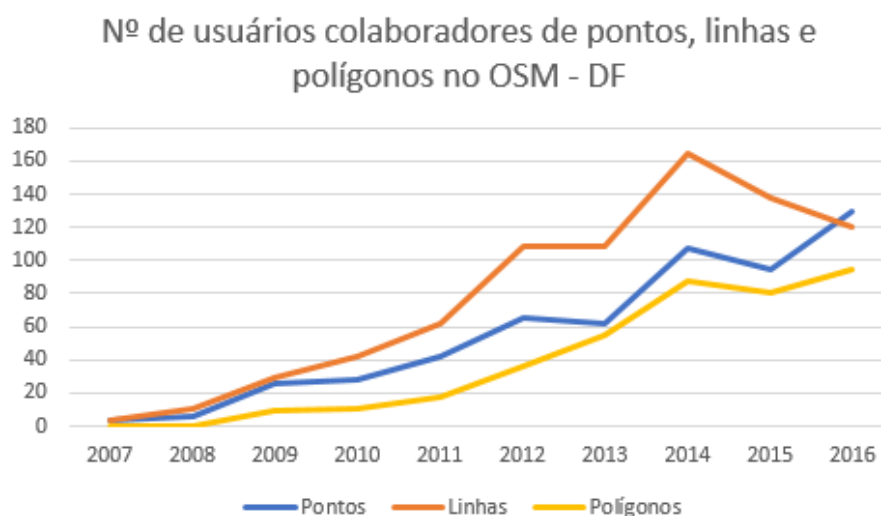


Figura 4.13: Nº de Usuários Colaboradores de Pontos, Linhas e Polígonos no OSM -DF.

A inserção de dados geográficos no Distrito Federal ocorreu de forma bastante curiosa. Na Figura 4.14 destacam-se os dois picos de inserção de linhas, ocorridos nos anos de 2011 e 2013. Além disso, percebe-se que a partir de 2013 ocorreu uma acentuada queda na inserção de dados no OSM - DF, o que poderia indicar de certa forma um abandono do projeto do OpenStreetMap no Distrito Federal a partir desse ano. Por outro lado, houve um aumento no número de polígonos inseridos na ferramenta, inclusive ultrapassando a quantidade de pontos ou linhas inseridos a partir do ano de 2015. Um dos possíveis

motivos para esse aumento foi que alguns lugares que antes eram modelados como pontos, tais quais escolas, prédios e restaurantes, passaram a ser modelados como polígonos.

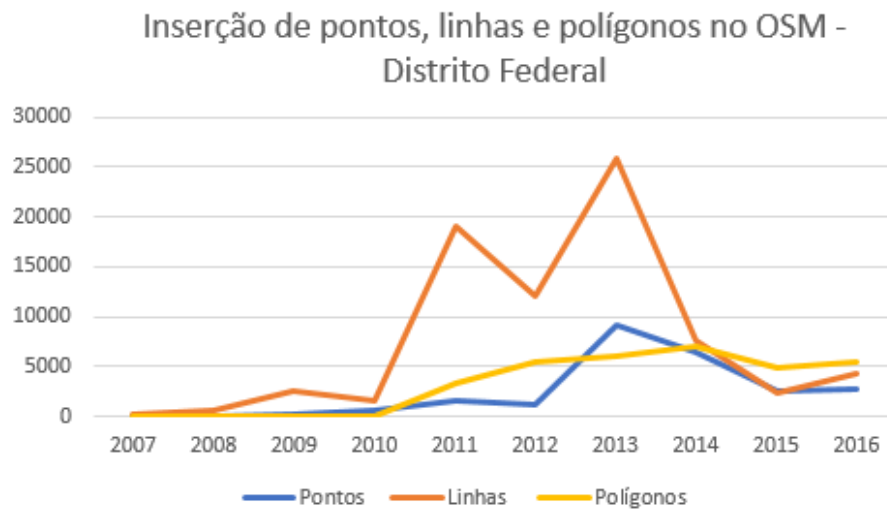


Figura 4.14: Inserção de Pontos, Linhas e Polígonos no OSM - DF.

Os dados do OpenStreetMap podem ser classificados de acordo com a *tag amenity*, utilizada para descrever informações úteis aos usuários do OSM. A Tabela 4.4 representa a distribuição de pontos no OSM - DF por categoria até o ano de 2016. Nessa tabela destaca-se o grande número de pontos sem identificação por meio da *tag amenity* (75,72% do total de pontos inseridos). Além disso, verifica-se o percentual de pontos do tipo Restaurante (4,08%), Escola (3,63%), Prédio (2,41%), Igreja (1,91%), Lanchonete (1,72%), Hospital (0,76%) e Outros (9,78%).

Tabela 4.4: Distribuição de Pontos no OSM - DF por Categoria até 2016.

Categoria	Participação
Sem Identificação	75,72%
Restaurante	4,08%
Escola	3,62%
Prédio	2,41%
Igreja	1,91%
Lanchonete	1,72%
Hospital	0,76%
Outros	9,78%

A Figura 4.15 representa a evolução dos tipos de pontos mais inseridos no Distrito Federal entre os anos de 2007 e 2016. Ao analisar a Figura 4.15, percebe-se que a partir de 2013 houve uma queda na inserção de quase todos os tipos de pontos verificados. Também a partir desse ano, a inserção do número de lanchonetes não teve uma queda tão

acentuada como os demais dados, mas estabilizou-se na faixa de 100 novas inclusões por ano.

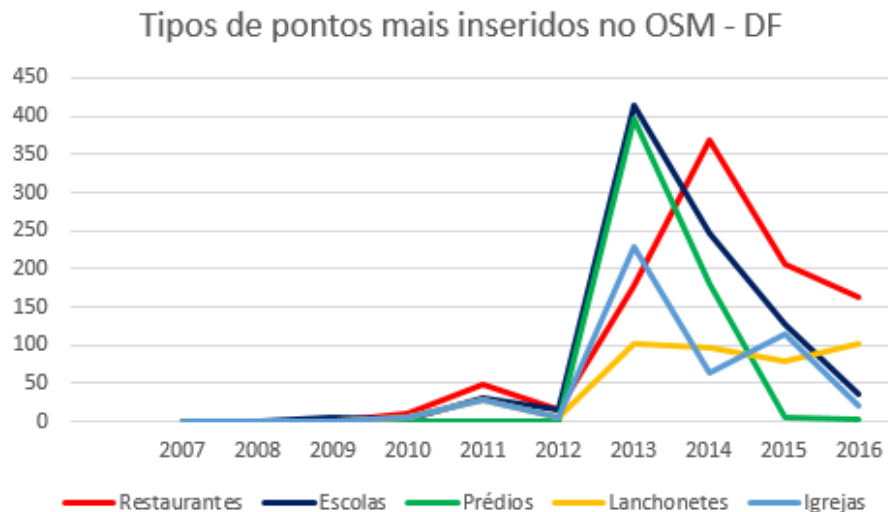


Figura 4.15: Tipos de Pontos Mais Inseridos no OSM - DF.

Alguns lugares, tais como escolas, hospitais ou prédios residenciais, podem ser modelados tanto como um ponto ou como um polígono, dependendo do grau de detalhamento que se deseja alcançar. Assim, se o usuário achar que a área do objeto deve ser considerada na representação, ele irá modelar o objeto como um polígono e caso contrário, irá modelar como ponto. A Figura 4.16 representa a evolução de pontos e polígonos com a *tag amenity* associada ao termo *school* (escola) na região correspondente ao bairro da Asa Norte, no Distrito Federal, para os anos de 2011, 2013, 2015 e 2016.

Ainda na Figura 4.16, percebe-se que ocorreu um salto na inclusão de pontos e polígonos na região da Asa Norte entre os anos de 2013 e 2015, mas não ocorreram mais tantas inserções no ano de 2016. Além disso, já no ano de 2011 percebe-se que há a presença de uma escola que foi modelada tanto como ponto como um polígono, indicando uma dupla representação.

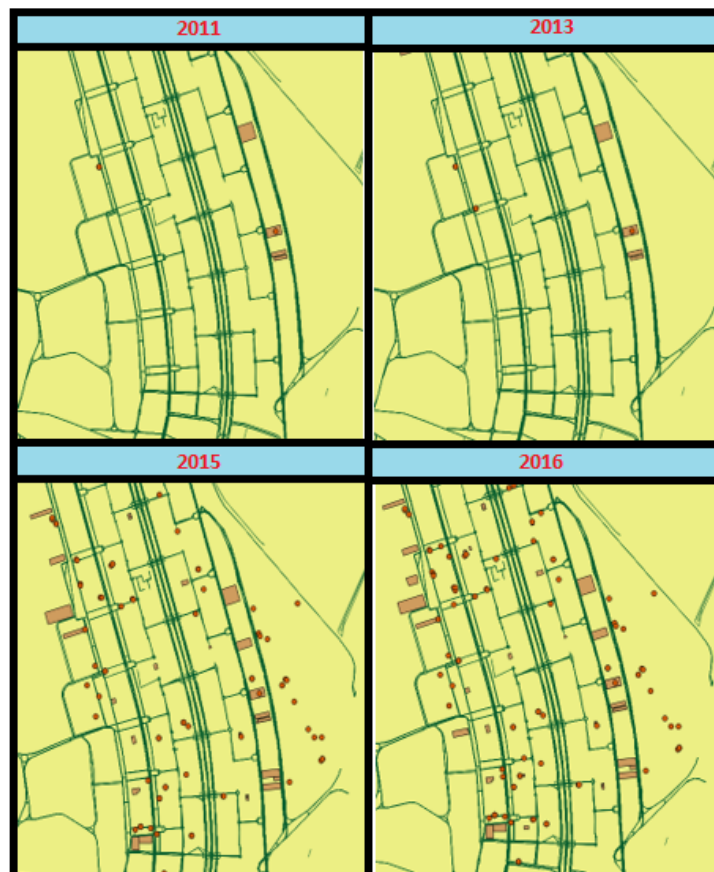


Figura 4.16: Evolução de Pontos e Polígonos do Tipo “Escola” na Região da Asa Norte.

Analisando-se a inserção de estradas no OSM - DF, percebe-se que as inclusões aconteceram de forma um tanto diferente do que ocorreu em nível nacional. A Tabela 4.5 representa a distribuição de estradas no Distrito Federal por categoria até o ano de 2016. Nessa tabela, apesar de ser verificada uma predominância de estradas próximas às áreas residenciais (32,53%), conforme também ocorreu com a inserção de estradas no Brasil, percebe-se que há uma ocorrência maior de vias secundárias (10,37%) e terciárias (17,04%) em relação às estradas primárias (9,75%).

Tabela 4.5: Distribuição de Estradas no OSM - DF por Categoria até 2016.

Categoria	Participação
Residential	32,53%
Tertiary	17,04%
Secondary	10,37%
Primary	9,75%
Unclassified	8,97%
Track	5,16%
Others	16,18%

A Figura 4.17 representa a evolução de estradas no Distrito Federal por categoria e para essa análise foi utilizada a *tag highway*, do mesmo modo que foi realizada a análise para as estradas brasileiras. Nessa figura pode-se perceber mais claramente os dois picos de inserção de dados do tipo linha, ocorridos nos anos de 2011 e 2013. No ano de 2011, cerca de 16 mil quilômetros de estradas foram acrescentados à ferramenta OpenStreetMap na região do Distrito Federal, e no ano de 2013 foram mais de 22 mil quilômetros inseridos.

Também na Figura 4.17 percebe-se que a partir do ano de 2013 ocorreu uma grande queda na inserção de dados na ferramenta OpenStretMap no DF, com praticamente o abandono do projeto pelos usuários que antes atuavam na região. À título de comparação com os anos de 2011 e 2013, no ano de 2016 foram incluídos menos de 2 mil quilômetros de estradas na região correspondente ao DF.

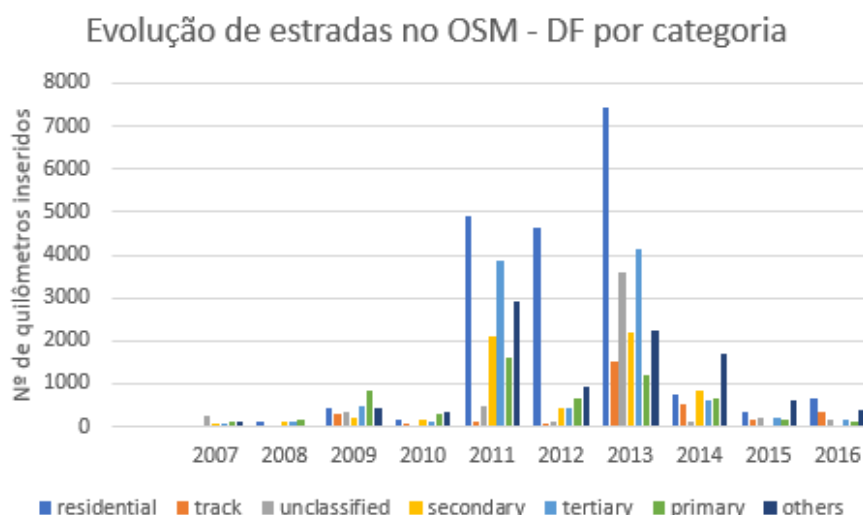


Figura 4.17: Evolução de Estradas no OSM - DF por Categoria até 2016.

A Figura 4.18 representa de que forma os dados geográficos do tipo linha foram inseridos na região correspondente ao Plano Piloto, área central do Distrito Federal. A partir dessa figura, percebe-se que no ano de 2008 os primeiros dados referentes à cidade de Brasília correspondiam ao contorno do Plano Piloto (no formato de avião) e ao contorno do lago Paranoá (no formato de curva). Com o passar dos anos, foram acrescentadas novas informações à base de dados, tais como construções e vias secundárias, além de ter se delimitado melhor os limites entre o lago e o restante da cidade.

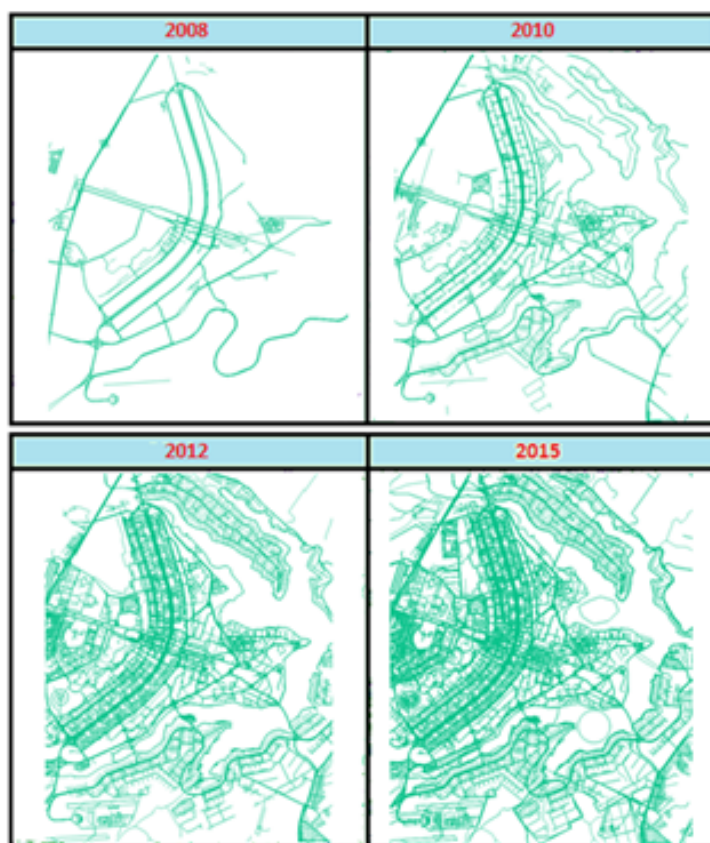


Figura 4.18: Evolução dos Dados do Tipo Linha no Distrito Federal.

Por meio das análises realizadas com a elaboração deste trabalho, pode-se perceber a relevância da produção de informação geográfica voluntária na atualidade. O próximo capítulo desta monografia é dedicado a apresentar as conclusões observadas a partir da realização deste trabalho, além de expor alguns trabalhos a serem realizados no futuro.

Capítulo 5

Conclusão e Trabalhos Futuros

Abordando a mesma temática desta monografia, foi escrito um outro trabalho no formato de artigo científico, o qual foi aceito na 12^a Conferência Ibérica de Sistemas e Tecnologias de Informação (CISTI - 2017) [20]. Por meio desses dois trabalhos, percebe-se que ferramentas como o OpenStreetMap, apesar de possuírem uma alta flexibilidade e permitir com que os próprios usuários realizem a edição de informações, contribuem para um índice elevado de heterogeneidade em relação aos dados inseridos, uma vez que o trabalho de contribuição colaborativa na maioria das vezes não ocorre de forma uniforme. Dessa maneira, algumas regiões tendem a concentrar mais dados do que outras. No caso do Brasil, verifica-se uma grande concentração de dados nas regiões Sul e Sudeste, enquanto há uma pouca concentração de dados na região Norte do país.

Em relação às estradas brasileiras foi possível verificar que as estradas mais importantes da malha viária foram as primeiras a serem inseridas, para que em um segundo momento os usuários pudessem focar em estradas próximas às áreas residenciais. Dessa forma, foi verificado um pico de inserções de estradas dos tipos *track* e *primary* no ano de 2009, quando a ferramenta ainda estava se consolidando no país. Porém, há uma predominância de vias residenciais a partir do ano de 2010.

Na região do Distrito Federal, foram verificados dois picos principais na inserção de dados, ocorridos nos anos de 2011 e 2013. Entretanto, após o ano de 2013, nota-se uma grande queda no número de inserções de dados na região correspondente ao DF e tal fato pode indicar um abandono dos usuários que antes atuavam nessa localidade.

Como trabalho futuro, sugere-se realizar outros tipos de análise, comparando-se a base de dados do OpenStreetMap no Brasil com a base de dados de *softwares* proprietários, a fim de verificar critérios de qualidade básicos, tais como completude e acurácia. Também deseja-se verificar a ocorrência de possíveis falhas nos dados presentes na ferramenta, como a quantidade de objetos sem nome ou verificação de objetos duplicados.

Referências

- [1] *Coordinate Reference Systems*. http://docs.qgis.org/2.0/en/docs/gentle_gis_introduction/coordinate_reference_systems.html. [Acessado em abril de 2017]. 15
- [2] *Estatísticas OSM - Brasil*. http://osmstats.neis-one.org/?item=countries_country=Brazil. [Acessado em novembro de 2016]. 35
- [3] *Neis*. <http://neis-one.org/>. [Acessado em novembro de 2016]. 34
- [4] *OpenStreetMap*. http://wiki.openstreetmap.org/wiki/About_OpenStreetMap. [Acessado em setembro de 2016]. 17
- [5] *Sistemas de Referência de Coordenadas*. <http://web.gps.caltech.edu/gislab/HowTo/ESRI%20-%20Map%20Projections.pdf>. [Acessado em abril de 2017]. 15
- [6] Borges, K. A. V.: *Modelagem de Dados Geográficos - Curso de Especialização em Geoprocessamento*. UFMG, 2002. ix, 6, 7, 8
- [7] Borges, K.A.V., C. A. Davis Jr e A. H. F. Laender: *OMT-G: An object-oriented data model for geographic applications*. Geoinformática, 3, 2002. 6
- [8] Del Fatto, V., V. Deufemia e L. Paolino: *Guaranteeing integrity constraints during interactive map editing with visual language parsing*. Digital Information Management, ICDIM '07. 2nd International Conference on, 1:19–24, 2007. 6, 7
- [9] Fonteles, A. S., Bouveret S. Gensel J.: *A generic architecture for spatial crowdsourcing*. IEEE Tenth International Conference on Research Challenges in Information Science (RCIS), páginas 1–10, 2016. ix, 16, 17
- [10] Frank, A. U.: *Spatial concepts, geometric data models and data structures*. Computers and Geosciences, 18. 5
- [11] Girres, J. F. e G. Touya: *Quality assessment of the French OpenStreetMap dataset*. Trans. GIS, 14:435–459, 2010. 21
- [12] Goodchild, M. F.: *Citizens as sensors: The world of volunteered geography*. GeoJournal, 69(4):211–221, 2007. 1, 13, 14
- [13] Güting, R. H.: *An introduction to spatial database systems*. VLDB Journal, 4:357–399, 1994. 5, 11

- [14] Haklay, M.: *How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets*. Environment and Planning B: Planning and Design, 37:682–703, 2010. 21
- [15] Haklay, M., Weber P.: *OpenStreetMap: User-generated street maps*. IEEE Pervasive Computing, 2008. ix, 18
- [16] Hammon, L., Hippner H.: *Crowdsourcing*. Business & Information Systems Engineering, 4:163–166, 2012. 16
- [17] Ivanov, V. Y., Vivoni E. R. Bras R. L. Entekhabi D.: *Catchment hydrologic response with a fully distributed triangulated irregular network model*. Water Resources Research, 40. 10
- [18] Kanapaka, R. R. N., Neelisete R. K: *A survey of tools for visualizing geo spatial data*. International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), páginas 22–27, 2015. ix, 13
- [19] Lisboa Filho, J.: *Projeto de Banco de Dados para Sistemas de Informação Geográfica*. VIII Escola de Informática da SBC Sul. Editora da UFSM, páginas 115–146, 2000. 5
- [20] Medeiros, G. F. B., Holanda M. T.: *Openstreetmap: An analysis of the evolution of geographic data in Brazil*. 12th Iberian Conference on Information Systems and Technologies (CISTI), 2017. 52
- [21] Monteiro, A. M., G. Camara, S.D. Fucks e M.S Carvalho: *Spatial analysis and gis: A primer*. National Institute for Space Research, 2001. ix, 8, 9, 11, 12
- [22] Navathe, S. B. e R. Elsmari: *Sistemas de Banco de Dados*. Editora Addison Wesley, 6ª edição, 2011. 4
- [23] Neis, P., Zielstra D. Zipf A.: *The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007-2011*. Future Internet, 4:1–21, 2012. 2, 22
- [24] Nielsen, J.: *The 90-9-1 rule for participation inequality in social media and online communities*. <https://www.nngroup.com/articles/participation-inequality/>, 2006. [Acessado em fevereiro de 2017]. 2
- [25] Pickles, J.: *A History of Spaces: Cartography reason, mapping and the geo-coded world*. Routledge, 2004. 1
- [26] Ramm, F., J. Topf e S. Chilton: *OpenStreetMap: Using and Enhancing the Free Map of the World*. UIT Cambridge, 2010. 1
- [27] RATIONAL, Software Corporation: *The Unified Language: notation guide, version 1.1*. 1997. 6
- [28] Rigaux, P., M. Scholl e A. Voisard: *Spatial Databases with Application to GIS*. Morgan Kaufmann Publishers, 2002. ix, 9, 10

- [29] Sehra, S. S., Singh J. Rai H. S: *A systematic study of openStreetMap data quality assessment*. Information Technology: New Generations (ITNG), 11th Internacional Conference on, páginas 377–381, 2014. 17, 20
- [30] Sui, D. Z.: *The wikification of GIS and its consequences: Or Angelina Jolie’s new tattoo and the future of GIS*. Computers, Environment and Urban Systems, 32(1):1–5, 2008. 1
- [31] Zhao, Y., Han Q.: *Spatial crowdsourcing: Current state and future directions*. IEEE Communications Magazine, 54(7):102–107, 2016. 16
- [32] Zielstra, D. e H. H. Hochmair: *Digital street data: Free versus proprietary*. GIM Int, 25:29–33, 2011. 21
- [33] Zielstra, D. e H. H. Hochmair: *Comparison of shortest path lengths for pedestrian routing in street networks using free and proprietary data*. Proceedings of Transportation Research Board - 91st Annual Meeting, páginas 22–26, 2012. 22
- [34] Zielstra, D. e A. Zipf: *A comparative study of proprietary geodata and volunteered geographic information for Germany*. Proceedings of 13th AGILE International Conference on Geographic Information Science, páginas 10–14, 2010. 21

Anexo I

Consultas SQL

1) Número de usuários colaboradores de pontos no ano de 2015:

```
select count(distinct tags->'osm_user') from planet_osm_point where tags->'osm_timestamp' like '2015%';
```

2) Número de pontos incluídos no ano de 2008:

```
select  
(select count(distinct osm_id) from planet_osm_point where tags->'osm_timestamp'  
like '2007%' or tags->'osm_timestamp' like '2008%')  
-  
(select count(distinct osm_id) from planet_osm_point where tags->'osm_timestamp'  
like '2007%') as Difference;
```

3) Extensão de estradas acrescentadas no OSM – Brasil por categoria no ano de 2007 (Resultado em quilômetros)

```
select line.highway, sum(ST_Length(ST_Transform(line.way,4326)::geography))/1000  
as roads_km from planet_osm_line line, planet_osm_polygon pol where pol.admin_level  
= '8' and st_contains (pol.way,line.way) and line.tags->'osm_timestamp' like '2007%'  
and line.highway is not null and line.tags->'osm_version' = '1' group by line.highway  
order by roads_km DESC;
```

4) Número de pontos inseridos nos anos de 2007 e 2008, agrupados por estado

```
select NM_ESTADO, count(distinct osm_id) from planet_osm_point p1, ufebrasil  
where (tags->'osm_timestamp' like '2007%' or tags->'osm_timestamp' like '2008%') and  
st_contains (geom,way)group by nm_estado;
```

5) Número de pontos inseridos entre os anos de 2007 e 2009, em municípios compreendidos pelo Grupo 1 (População menor que 10.000 habitantes no ano de 2010)

```
select count(distinct osm_id) from municipios_2010, planet_osm_point where st_intersects (geom,way) and (tags->'osm_timestamp' like '2007%' or tags->'osm_timestamp' like '2008%' or tags->'osm_timestamp' like '2009%') and populacao <10000;
```

6) Número de pontos inseridos no Distrito Federal no ano de 2008, separados por categoria

```
select p1.amenity, count(distinct p1.osm_id) from planet_osm_point p1, planet_osm_polygon p2 where p2.admin_level = '4' and (p2.name = 'D.F.') and st_contains(p2.way, p1.way) and (p1.tags->'osm_timestamp' like '2008%') group by p1.amenity order by count DESC;
```